

Learn-and-Adapt Stochastic Dual Gradients for Network Resource Allocation

Tianyi Chen, Qing Ling, and Georgios B. Giannakis

Abstract—Network resource allocation shows revived popularity in the era of data deluge and information explosion. Existing stochastic optimization approaches fall short in attaining a desirable cost-delay tradeoff. Recognizing the central role of Lagrange multipliers in network resource allocation, a novel learn-and-adapt stochastic dual gradient (LA-SDG) method is developed in this paper to learn the empirical optimal Lagrange multiplier from historical data, and adapt to the upcoming resource allocation strategy. Remarkably, it only requires one more sample (gradient) evaluation than the celebrated stochastic dual gradient (SDG) method. LA-SDG can be interpreted as a foresighted learning approach with *an eye on the future*, or, a modified heavy-ball approach from an optimization viewpoint. It is established - both theoretically and empirically - that LA-SDG markedly improves the cost-delay tradeoff over state-of-the-art allocation schemes.

Index Terms—First-order method, stochastic approximation, statistical learning, network resource allocation.

I. INTRODUCTION

In the era of big data analytics, cloud computing and Internet of Things, the growing demand of massive data processing challenges existing network resource allocation approaches. Huge volumes of data acquired online with distributed sensors in the presence of operational uncertainties caused by, e.g., renewable energy sources, call for scalable and adaptive network control schemes. Scalability of a desirable approach refers to low complexity and amenability to distributed implementation, while adaptivity implies capability of online adjustment to dynamic environments.

Allocation of network resources can be traced back to the seminal work of [1]. Since then, popular allocation algorithms operating in the dual domain are first-order methods based on dual gradient ascent, either deterministic [2] or stochastic [3], [4]. Due to the simplicity in computation and implementation, these approaches have attracted a great deal of recent interest and have been successfully applied to cloud, transportation and power grid networks; see, e.g., [5]–[8]. However, their major limitation is slow convergence, which results in high network delay. Here, the delay can be viewed as workload queuing time in a cloud network, traffic congestion in a transportation network, or energy level of batteries in a power network. To address this issue, recent attempts aim at accelerating first- and second-order optimization algorithms [9]–[12]. Specifically,

momentum-based accelerations over first-order methods were investigated using Nesterov [9], or, heavy-ball iterations [10]. Though these approaches work well in static settings, their performance degrades with online scheduling, as evidenced by the increase in accumulated steady-state error [13]. On the other hand, second-order methods such as decentralized quasi-Newton approach and its dynamic variant developed in [11] and [12], incur high overhead to compute and communicate the decentralized Hessian approximations.

Capturing prices of resources, Lagrange multipliers play a central role in stochastic resource allocation algorithms [14]. Given abundant historical data in an online optimization setting, a natural question arises: *Is it possible to learn the optimal prices from past data, so as to improve the quality of online resource allocation strategies?* The rationale here is that past data contains statistics of network states, and learning from them can aid coping with the stochasticity of future resource allocation. A recent work attempting to address this question is [15], which considers resource allocation with a *finite* number of possible network states and allocation actions. The learning procedure, however, involves constructing a histogram to estimate the underlying distribution of the network states and explicitly solves an empirical dual problem. While constructing a histogram is feasible for a probability distribution with finite support, quantization errors and prohibitively high complexity are inevitable for a continuous distribution with infinite support.

In this context, the present paper aims to design a novel online resource allocation algorithm that leverages online learning from historical data for stochastic optimization of the ensuing allocation stage. The resultant approach, which we term “learn-and-adapt” stochastic dual gradient (LA-SDG) method, only doubles computational complexity of the classic stochastic dual gradient (SDG) method. With this minimal cost, LA-SDG mitigates steady-state oscillation, which is common in stochastic first-order acceleration methods [10], [13], while avoiding computation of the Hessian approximations present in the second-order methods [11], [12]. Specifically, LA-SDG only requires one more past sample to compute an extra stochastic dual gradient, in contrast to constructing costly histograms and solving the resultant large-scale problem [15].

The main contributions of this paper are summarized next.

- c1) Targeting a low-complexity online solution, LA-SDG only takes an additional dual gradient step relative to the classic SDG iteration. This additional dual gradient step plays the role of adapting the future resource allocation strategy through learning from historical data. Meanwhile, LA-SDG is linked with the stochastic heavy-ball method,

Work in this paper was supported by NSF 1509040, 1508993, 1509005, NSF China 61573331, NSF Anhui 1608085QF130, and cas-xda06011203.

T. Chen and G. B. Giannakis are with the Department of Electrical and Computer Engineering and the Digital Technology Center, University of Minnesota, Minneapolis, MN 55455 USA. Emails: {chen3827, georgios}@umn.edu

Q. Ling is with the Department of Automation, University of Science and Technology of China, Hefei, Anhui 230026, China. Email: qinling@mail.ustc.edu.cn

nicely inheriting its fast convergence in the initial stage, while reducing its steady-state oscillation.

- c2) We analytically establish that the novel LA-SDG approach, parameterized by a positive constant μ , yields a very attractive cost-delay tradeoff $[\mu, \log^2(\mu)/\sqrt{\mu}]$, which is better than the standard tradeoff $[\mu, 1/\mu]$ of the classic SDG method [4]. Numerical tests further corroborate the performance gain of LA-SDG over existing resource allocation schemes.

Outline. The rest of the paper is organized as follows. The stochastic network resource allocation problem is formulated in Section II, along with the standard SDG approach introduced in Section III. The LA-SDG method is the subject of Section IV. Performance analysis of LA-SDG is carried out in Section V. Numerical tests are provided in Section VI, followed by concluding remarks in Section VII.

Notation. \mathbb{E} denotes the expectation operator, \mathbb{P} stands for probability; $(\cdot)^\top$ stands for vector and matrix transposition, and $\|\mathbf{x}\|$ denotes the ℓ_2 -norm of a vector \mathbf{x} . Inequalities for vectors, e.g., $\mathbf{x} > \mathbf{0}$, are defined entry-wise. The positive projection operator is defined as $[a]^+ := \max\{a, 0\}$, also entry-wise.

II. NETWORK RESOURCE ALLOCATION

In this section, we start with a generic network model and its resource allocation task in Section II-A, and then introduce a specific example of resource allocation in cloud networks in Section II-B. The proposed approach is applicable to more general network resource allocation tasks such as geographical load balancing in cloud networks [5], traffic control in transportation networks [7], and energy management in power networks [8].

A. A unified resource allocation model

Consider discrete time $t \in \mathbb{N}$, and a network represented as a directed graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ with nodes $\mathcal{I} := \{1, \dots, I\}$ and edges $\mathcal{E} := \{1, \dots, E\}$. Collect the workloads across edges $e = (i, j) \in \mathcal{E}$ in a resource allocation vector $\mathbf{x}_t \in \mathbb{R}^E$. The $I \times E$ node-incidence matrix is formed with the (i, e) -th entry

$$\mathbf{A}_{(i,e)} = \begin{cases} 1, & \text{if link } e \text{ enters node } i \\ -1, & \text{if link } e \text{ leaves node } i \\ 0, & \text{else.} \end{cases} \quad (1)$$

We assume that each row of \mathbf{A} has at least one -1 entry, and each column of \mathbf{A} has at most one -1 entry, meaning that each node has at least one outgoing link, and each link has at most one source node. With $\mathbf{c}_t \in \mathbb{R}_+^I$ collecting the random workload arrival rates of all nodes per slot t , the aggregate (endogenous plus exogenous) workloads of all nodes are $\mathbf{A}\mathbf{x}_t + \mathbf{c}_t$. If the i -th entry of $\mathbf{A}\mathbf{x}_t + \mathbf{c}_t$ is positive, there is service residual queued at node i ; otherwise, node i over-serves the current arrival. With a workload queue per node, the queue length vector $\mathbf{q}_t := [q_t^1, \dots, q_t^I]^\top \in \mathbb{R}_+^I$ obeys the recursion

$$\mathbf{q}_{t+1} = [\mathbf{q}_t + \mathbf{A}\mathbf{x}_t + \mathbf{c}_t]^+, \quad \forall t. \quad (2)$$

Defining $\Psi_t(\mathbf{x}_t) := \Psi(\mathbf{x}_t; \phi_t)$ as the aggregate network cost parameterized by the random vector ϕ_t , the local cost per node i is $\Psi_t^i(\mathbf{x}_t) := \Psi^i(\mathbf{x}_t; \phi_t^i)$, and $\Psi_t(\mathbf{x}_t) = \sum_{i \in \mathcal{I}} \Psi_t^i(\mathbf{x}_t)$. The model here is quite general. The duration of time slots can

vary from (micro-)seconds in cloud networks, minutes in road networks, to even hours in power networks; the nodes can present the distributed front-end mapping nodes and back-end data centers in cloud networks, intersections in traffic networks, or, buses and substations in power networks; the links can model wireless/wireline channels, traffic lanes, and power transmission lines; while the resource vector \mathbf{x}_t can include the size of data workloads, the number of vehicles, or the amount of energy. Vector \mathbf{q}_t can represent user requests in data queues, cars waiting at road intersections, and amounts of energy in batteries.

Concatenating the random parameters into a random state vector $\mathbf{s}_t := [\phi_t^\top, \mathbf{c}_t^\top]^\top$, the resource allocation task is to determine the allocation \mathbf{x}_t in response to the observed (realization) \mathbf{s}_t “on the fly,” so as to minimize the *long-term average* network cost subject to queue stability at each node, and operation feasibility at each link. Concretely, we have

$$\Psi^* := \min_{\{\mathbf{x}_t, \mathbf{q}_t\}} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t)] \quad (3a)$$

$$\text{s.t. } \mathbf{q}_{t+1} = [\mathbf{q}_t + \mathbf{A}\mathbf{x}_t + \mathbf{c}_t]^+, \quad \forall t \quad (3b)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{q}_t] < \infty \quad (3c)$$

$$\mathbf{x}_t \in \mathcal{X} := \{\mathbf{x}_t \mid \mathbf{0} \leq \mathbf{x}_t \leq \bar{\mathbf{x}}\}, \quad \forall t \quad (3d)$$

where Ψ^* is the optimal objective of problem (3) given all the future information; \mathbb{E} is taken over \mathbf{s}_t as well as possible randomness of optimization variables \mathbf{x}_t and \mathbf{q}_t ; constraints (3c) ensure the queue stability¹; and (3d) constrains the instantaneous allocation variables to be within a time-invariant box constraint set \mathcal{X} , which is specified by, e.g., link capacities, or, server/generator capacities.

The queue dynamics in (3b) couple the optimization variables over an infinite time horizon, which implies that the decision variable at current slot will have effect on all the future decisions. Therefore, finding an optimal solution to (3) calls for dynamic programming [16], which is known to suffer from the “curse of dimensionality” and intractability in an online setting. In Section III-A, we will circumvent this obstacle by relaxing (3b)-(3c) to limiting average constraints, and employing dual decomposition techniques.

B. Motivating setup

The geographic load balancing task in cloud networks [5], [17], [18] takes the form of (3) with $\mathcal{J} = \{1, \dots, J\}$ being the set of mapping nodes (e.g., DNS servers), $\mathcal{K} = \{1, \dots, K\}$ data centers, and $\mathcal{E} := \{(j, k), \forall j \in \mathcal{J}, k \in \mathcal{K}\}$ links connecting mapping nodes with data centers. Per time t , each mapping node j collects the amount of user data requests c_t^j , and forwards the amount x_t^{jk} to data center k constrained by the bandwidth availability. Each data center k schedules workload processing y_t^k according to its resource availability. To match the definition of node-incidence matrix in (1), we assume a virtual outgoing node from each data center k , so that y_t^k is the amount of resources on this outgoing link. The bandwidth limit of link

¹Here we focus on the strong stability given by [4, Definition 2.7], which requires the time-average expected queue length to be finite.

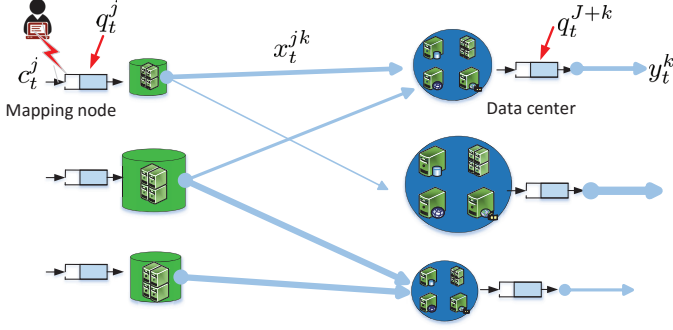


Fig. 1. A diagram of online geographical load balancing. Per time t , mapping node j has an exogenous workload c_t^j plus that stored in the queue q_t^j , and schedules workload x_t^{jk} to data center k . Data center k serves an amount of workload y_t^k out of the assigned x_t^{jk} as well as that stored in the queue q_t^{J+k} . The thickness of each edge is proportional to its capacity.

(j, k) is \bar{x}^{jk} , and the resource capability of data center k is \bar{y}^k . In this case, we have $\mathbf{x}_t := [x_t^{11}, \dots, x_t^{JK}, y_t^1, \dots, y_t^K]^\top$, $\mathbf{c}_t := [c_t^1, \dots, c_t^J, 0, \dots, 0]^\top$, and $\bar{\mathbf{x}} := [\bar{x}^{11}, \dots, \bar{x}^{JK}, \bar{y}^1, \dots, \bar{y}^K]^\top$. At each mapping node and data center, undistributed or unprocessed workloads are buffered in queues obeying (3b); see Fig. 1 for a system diagram.

Performance is characterized by the aggregate cost of power consumed at the data centers plus the bandwidth costs at the mapping nodes, namely

$$\Psi_t(\mathbf{x}_t) := \sum_{k \in \mathcal{K}} \underbrace{\Psi_t^k(y_t^k)}_{\text{power cost}} + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \underbrace{\Psi_t^{jk}(x_t^{jk})}_{\text{bandwidth cost}}. \quad (4)$$

The power cost $\Psi_t^k(y_t^k) := \Psi(y_t^k; \phi_t^k)$, parameterized by the random vector ϕ_t^k , captures the local marginal price and the renewable generation at data center k during time period t . The bandwidth cost $\Psi_t^{jk}(x_t^{jk}) := \Psi^{jk}(x_t^{jk}; \phi_t^{jk})$, parameterized by the random vector ϕ_t^{jk} , characterizes the heterogeneous cost of data transmission due to spatio-temporal differences. With the goal of minimizing the time-average of (4), geographical load balancing fits the formulation in (3).

III. ONLINE NETWORK MANAGEMENT VIA SDG

In this section, the dynamic problem (3) is reformulated to a tractable form, and classical stochastic dual gradient (SDG) approach is revisited, along with a brief discussion of its online performance.

A. Problem reformulation

Recall in Section II-A that the main challenge of solving (3) resides in time-coupling constraints and unknown distribution of the underlying random processes. Regarding the first hurdle, combining (3b) with (3c), it can be shown that in the long term, workload arrival and departure rates must satisfy the following necessary condition [4, Theorem 2.8]

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{A}\mathbf{x}_t + \mathbf{c}_t] \leq \mathbf{0} \quad (5)$$

given that the initial queue length is finite, i.e., $\|\mathbf{q}_1\| \leq \infty$. In other words, on average all buffered delay-tolerant workloads should be served. Using (5), a relaxed version of (3) is

$$\tilde{\Psi}^* := \min_{\{\mathbf{x}_t, \forall t\}} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t)] \quad \text{s.t. (3d) and (5)} \quad (6)$$

where $\tilde{\Psi}^*$ is the optimal objective for the relaxed problem (6).

Compared to (3), problem (6) eliminates the time coupling across variables $\{\mathbf{q}_t, \forall t\}$ by replacing (3b) and (3c) with (5). Since (6) is a relaxed version of (3) with the optimal objective $\tilde{\Psi}^* \leq \Psi^*$, if one solves (6) instead of (3), it will be prudent to derive an optimality bound on Ψ^* , provided that the sequence of solutions $\{\mathbf{x}_t, \forall t\}$ obtained by solving (6) is feasible for the relaxed constraints (3b) and (3c). Regarding the relaxed problem (6), using arguments similar to those in [4, Theorem 4.5], it can be shown that if the random state \mathbf{s}_t is independent and identically distributed (i.i.d.) over time t , there exists a *stationary* control policy $\chi^*(\cdot)$, which is a pure (possibly randomized) function of the realization of random state \mathbf{s}_t (or the *observed* state \mathbf{s}_t); i.e., it satisfies (3d), as well as guarantees that $\mathbb{E}[\Psi_t(\chi^*(\mathbf{s}_t))] = \tilde{\Psi}^*$ and $\mathbb{E}[\mathbf{A}\chi^*(\mathbf{s}_t) + \mathbf{c}_t] \leq \mathbf{0}$. As the optimal policy $\chi^*(\cdot)$ is time invariant, it implies that the *dynamic* problem (6) is equivalent to the following time-invariant *ensemble* program

$$\tilde{\Psi}^* := \min_{\chi(\cdot)} \mathbb{E}[\Psi(\chi(\mathbf{s}_t); \mathbf{s}_t)] \quad (7a)$$

$$\text{s.t. } \mathbb{E}[\mathbf{A}\chi(\mathbf{s}_t) + \mathbf{c}(\mathbf{s}_t)] \leq \mathbf{0} \quad (7b)$$

$$\chi(\mathbf{s}_t) \in \mathcal{X}, \quad \forall \mathbf{s}_t \in \mathcal{S} \quad (7c)$$

where $\chi(\mathbf{s}_t) := \mathbf{x}_t$, $\mathbf{c}(\mathbf{s}_t) = \mathbf{c}_t$, and $\Psi(\chi(\mathbf{s}_t); \mathbf{s}_t) := \Psi_t(\mathbf{x}_t)$; set \mathcal{S} is the sample space of \mathbf{s}_t , and the constraint (7c) holds almost surely. Observe that the index t in (7) can be dropped, since the expectation is taken over the distribution of random variable \mathbf{s}_t , which is time-invariant. Leveraging the equivalent form (7), the remaining task boils down to finding the optimal policy that achieves the minimal objective in (7a) and obeys the constraints (7b) and (7c).² Note that the optimization in (7) is with respect to a stationary policy $\chi(\cdot)$, which is an infinite dimensional problem in the primal domain. However, there is a finite number of expected constraints [cf. (7b)]. Thus, the dual problem contains a finite number of variables, hinting to the effect that solving (7) is tractable in the dual domain [16], [19], [20].

B. Lagrange dual and optimal policy

With $\boldsymbol{\lambda} \in \mathbb{R}_+^I$ denoting the Lagrange multipliers associated with (7b), the Lagrangian of (7) is

$$\mathcal{L}(\chi, \boldsymbol{\lambda}) := \mathbb{E}[\mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda})] \quad (8)$$

where the instantaneous Lagrangian is

$$\mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda}) := \Psi_t(\mathbf{x}_t) + \boldsymbol{\lambda}^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \quad (9)$$

and constraint (7c) remains implicit. Notice that the instantaneous objective $\Psi_t(\mathbf{x}_t)$ and the instantaneous constraint

²Though there may exist other time-dependent policies that generate the optimal solution to (6), our attention is restricted to the one that purely depends on the observed state $\mathbf{s} \in \mathcal{S}$, which can be time-independent [4, Theorem 4.5].

$\mathbf{A}\mathbf{x}_t + \mathbf{c}_t$ are both parameterized by the observed state $\mathbf{s}_t := [\phi_t^\top, \mathbf{c}_t^\top]^\top$ at time t ; thus, the instantaneous Lagrangian can be written as $\mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda}) = \mathcal{L}(\boldsymbol{\chi}(\mathbf{s}_t), \boldsymbol{\lambda}; \mathbf{s}_t)$, so that $\mathcal{L}(\boldsymbol{\chi}, \boldsymbol{\lambda}) := \mathbb{E}[\mathcal{L}(\boldsymbol{\chi}(\mathbf{s}_t), \boldsymbol{\lambda}; \mathbf{s}_t)]$.

Correspondingly, the Lagrange dual function is defined as the minimum of the Lagrangian over the all feasible primal variables [21], given by

$$\begin{aligned} \mathcal{D}(\boldsymbol{\lambda}) &:= \min_{\{\boldsymbol{\chi}(\mathbf{s}_t) \in \mathcal{X}, \forall \mathbf{s}_t \in \mathcal{S}\}} \mathcal{L}(\boldsymbol{\chi}, \boldsymbol{\lambda}) \\ &= \min_{\{\boldsymbol{\chi}(\mathbf{s}_t) \in \mathcal{X}, \forall \mathbf{s}_t \in \mathcal{S}\}} \mathbb{E}[\mathcal{L}(\boldsymbol{\chi}(\mathbf{s}_t), \boldsymbol{\lambda}; \mathbf{s}_t)]. \end{aligned} \quad (10a)$$

Due to the linearity of the expectation operator and the separability of the instantaneous Lagrangian defined by the realization \mathbf{s}_t , we can interchange the minimization and the expectation, and re-write the dual function in the following form

$$\mathcal{D}(\boldsymbol{\lambda}) = \mathbb{E} \left[\min_{\boldsymbol{\chi}(\mathbf{s}_t) \in \mathcal{X}} \mathcal{L}(\boldsymbol{\chi}(\mathbf{s}_t), \boldsymbol{\lambda}; \mathbf{s}_t) \right] = \mathbb{E} \left[\min_{\mathbf{x}_t \in \mathcal{X}} \mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda}) \right]. \quad (10b)$$

Likewise, for the instantaneous dual function $\mathcal{D}_t(\boldsymbol{\lambda}) = \mathcal{D}(\boldsymbol{\lambda}; \mathbf{s}_t) := \min_{\mathbf{x}_t \in \mathcal{X}} \mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda})$, the dual problem of (7) is

$$\max_{\boldsymbol{\lambda} \geq 0} \mathcal{D}(\boldsymbol{\lambda}) := \mathbb{E}[\mathcal{D}_t(\boldsymbol{\lambda})]. \quad (11)$$

In accordance with the ensemble primal problem (7), we will henceforth refer to (11) as the *ensemble* dual problem.

If the optimal Lagrange multiplier $\boldsymbol{\lambda}^*$ associated with (7b) were known, then optimizing (7) and consequently (6) would be equivalent to minimizing the Lagrangian $\mathcal{L}(\boldsymbol{\chi}, \boldsymbol{\lambda}^*)$ or infinitely many instantaneous $\{\mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda}^*)\}$, over the set \mathcal{X} [22, Prop. 3.3.4] and [16]. We restate this assertion as follows.

Proposition 1. *Consider the optimization problem in (7). Given a realization \mathbf{s}_t , and the optimal Lagrange multiplier $\boldsymbol{\lambda}^*$ associated with the constraints (7b), the optimal instantaneous resource allocation decision is*

$$\mathbf{x}_t^* = \boldsymbol{\chi}^*(\mathbf{s}_t) \in \arg \min_{\boldsymbol{\chi}(\mathbf{s}_t) \in \mathcal{X}} \mathcal{L}(\mathbf{x}_t, \boldsymbol{\lambda}^*; \mathbf{s}_t) \quad (12)$$

where \in accounts for possibly multiple minimizers of \mathcal{L}_t .

When the realizations $\{\mathbf{s}_t\}$ are obtained sequentially, one can generate a sequence of optimal solutions $\{\mathbf{x}_t^*\}$ correspondingly for the dynamic problem (6). To obtain the optimal allocation in (12) however, $\boldsymbol{\lambda}^*$ must be known. This fact motivates our novel “learn-and-adapt” stochastic dual gradient (LA-SDG) method in Section IV. To this end, we will first outline the celebrated stochastic dual gradient iteration (a.k.a. Lyapunov optimization).

C. Revisiting stochastic dual gradient

To solve (11), a standard gradient iteration involves sequentially taking expectations over the distribution of \mathbf{s}_t to compute the gradient. This is challenging because the distribution of \mathbf{s}_t is usually unknown in practice. Even if the joint probability distribution functions were available, finding the expectations is not scalable as the dimensionality of \mathbf{s}_t grows.

A common remedy to this challenge is stochastic approximation [4], [23], which corresponds to the following SDG iteration

$$\boldsymbol{\lambda}_{t+1} = [\boldsymbol{\lambda}_t + \mu \nabla \mathcal{D}_t(\boldsymbol{\lambda}_t)]^+, \forall t \quad (13a)$$

where μ is a positive (and typically pre-selected constant) stepsize. The stochastic gradient $\nabla \mathcal{D}_t(\boldsymbol{\lambda}_t) = \mathbf{A}\mathbf{x}_t + \mathbf{c}_t$ is an unbiased estimate of the true gradient; that is, $\mathbb{E}[\nabla \mathcal{D}_t(\boldsymbol{\lambda}_t)] = \nabla \mathcal{D}(\boldsymbol{\lambda}_t)$. Hence, the primal \mathbf{x}_t can be found by solving the following instantaneous sub-problems, one per t

$$\mathbf{x}_t \in \arg \min_{\mathbf{x}_t \in \mathcal{X}} \mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\lambda}_t). \quad (13b)$$

The iterate $\boldsymbol{\lambda}_{t+1}$ in (13a) depends only on the probability distribution of \mathbf{s}_t through the stochastic gradient $\nabla \mathcal{D}_t(\boldsymbol{\lambda}_t)$. Consequently, the process $\{\boldsymbol{\lambda}_t\}$ is Markov with invariant transition probability when \mathbf{s}_t is stationary. An interesting observation is that since $\nabla \mathcal{D}_t(\boldsymbol{\lambda}_t) := \mathbf{A}\mathbf{x}_t + \mathbf{c}_t$, the dual iteration can be written as [cf. (13a)]

$$\boldsymbol{\lambda}_{t+1}/\mu = [\boldsymbol{\lambda}_t/\mu + \mathbf{A}\mathbf{x}_t + \mathbf{c}_t]^+, \forall t \quad (14)$$

which coincides with (3b) for $\boldsymbol{\lambda}_t/\mu = \mathbf{q}_t$; see also [4], [14], [17] for the virtual queue interpretation of this parallelism.

Thanks to its low complexity and robustness to non-stationary scenarios, SDG is widely used in various areas, including adaptive signal processing [24], stochastic network optimization [4], [14], [15], and energy management in power grids [8], [17]. For network management in particular, this iteration entails a cost-delay tradeoff as summarized next; see e.g., [4].

Proposition 2. *If Ψ^* is the optimal cost in (3) under any feasible control policy with the state distribution available, and if the SDG recursions in (13a)-(13b) achieve an $\mathcal{O}(\mu)$ -optimal solution in the sense that*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t(\boldsymbol{\lambda}_t))] \leq \Psi^* + \mathcal{O}(\mu) \quad (15a)$$

where $\mathbf{x}_t(\boldsymbol{\lambda}_t)$ denotes the decisions obtained from (13b), then it necessarily incurs a steady-state queue length $\mathcal{O}(1/\mu)$, namely

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{q}_t] = \mathcal{O}\left(\frac{1}{\mu}\right). \quad (15b)$$

Proposition 2 asserts that SDG will asymptotically yield an $\mathcal{O}(\mu)$ -optimal solution, and have steady-state queue length \mathbf{q}_∞ inversely proportional to μ . Intuitively, SDG iteration (13a) with a constant stepsize μ will converge to a neighborhood of the optimum $\boldsymbol{\lambda}^*$ [24]. Under mild conditions, the optimal multiplier is bounded, i.e., $\boldsymbol{\lambda}^* = \mathcal{O}(1)$, so that the steady-state queue length $\mathbf{q}_\infty = \boldsymbol{\lambda}_\infty/\mu$ naturally scales with $\mathcal{O}(1/\mu)$; see (14). As a consequence however, to achieve the near optimality (sufficiently small μ), SDG suffers large average queue lengths, thus undesired average delay by Little’s law [4]. To overcome this limitation, we will develop next an *online* approach, which can improve SDG’s cost-delay tradeoff, while still preserving its affordable complexity and adaptability.

IV. LEARN-AND-ADAPT SDG

Our main approach is derived in this section, by nicely leveraging both learning and optimization tools. Its decentralized implementation is also developed.

Algorithm 1 LA-SDG for Stochastic Network Optimization

- 1: **Initialize:** dual iterate λ_1 , empirical dual iterate $\hat{\lambda}_1$, queue length \mathbf{q}_1 , control variable $\theta = \sqrt{\mu} \log^2(\mu) \cdot \mathbf{1}$, and proper stepsizes μ and $\{\eta_t, \forall t\}$.
 - 2: **for** $t = 1, 2, \dots$ **do**
 - 3: **Resource allocation (1st gradient):**
 - 4: Construct the effective dual variable via (17b), observe the current state \mathbf{s}_t , and obtain resource allocation $\mathbf{x}_t(\gamma_t)$ by minimizing online Lagrangian (17a).
 - 5: Update the instantaneous queue length \mathbf{q}_{t+1} via

$$\mathbf{q}_{t+1} = [\mathbf{q}_t + (\mathbf{A}\mathbf{x}_t(\gamma_t) + \mathbf{c}_t)]^+, \forall t. \quad (16)$$
 - 6: **Sample recourse (2nd gradient):**
 - 7: Obtain variable $\mathbf{x}_t(\hat{\lambda}_t)$ by solving online Lagrangian minimization with sample \mathbf{s}_t via (18b).
 - 8: Update the empirical dual variable $\hat{\lambda}_{t+1}$ via (18a).
 - 9: **end for**
-

A. LA-SDG as a foresighted learning scheme

The intuition behind our learn-and-adapt stochastic dual gradient (LA-SDG) approach is to incrementally learn network state statistics from observed data while adapting resource allocation driven by the learning process. A key element of LA-SDG could be termed as “foresighted” learning because instead of myopically learning the exact optimal argument from empirical data, LA-SDG maintains the capability to hedge against the risk of “future non-stationarities.”

The proposed LA-SDG is summarized in Algorithm 1. It involves the queue length \mathbf{q}_t and an empirical dual variable $\hat{\lambda}_t$, along with a bias-control variable θ to ensure that LA-SDG will attain near optimality in the steady state [cf. Theorems 2 and 3]. At each time slot t , LA-SDG obtains two stochastic gradients using the current \mathbf{s}_t : One for online resource allocation, and another one for sample learning/recourse. For the first gradient (lines 3-5), contrary to SDG that relies on the stochastic multiplier estimate λ_t [cf. (13b)], LA-SDG minimizes the instantaneous Lagrangian

$$\mathbf{x}_t(\gamma_t) \in \arg \min_{\mathbf{x}_t \in \mathcal{X}} \mathcal{L}_t(\mathbf{x}_t, \gamma_t) \quad (17a)$$

which depends on what we term *effective* multiplier γ_t , given by

$$\underbrace{\gamma_t}_{\text{effective multiplier}} = \underbrace{\hat{\lambda}_t}_{\text{statistical learning}} + \underbrace{\mu \mathbf{q}_t - \theta}_{\text{online adaptation}}, \forall t. \quad (17b)$$

Variable γ_t also captures the effective price, which is a linear combination of the empirical $\hat{\lambda}_t$ and the queue length \mathbf{q}_t , where the control variable μ tunes the weights of these two factors, and θ controls the bias of γ_t in the steady state [15]. As a single pass of SDG “wastes” valuable online samples, LA-SDG resolves this limitation in a learning step by evaluating a second gradient (lines 6-8); that is, LA-SDG simply finds the stochastic gradient of (11) at the previous empirical dual variable $\hat{\lambda}_t$, and implements a gradient ascent update as

$$\hat{\lambda}_{t+1} = [\hat{\lambda}_t + \eta_t (\mathbf{A}\mathbf{x}_t(\hat{\lambda}_t) + \mathbf{c}_t)]^+, \forall t \quad (18a)$$

where η_t is a proper diminishing stepsize, and the “virtual” allocation $\mathbf{x}_t(\hat{\lambda}_t)$ can be found by solving

$$\mathbf{x}_t(\hat{\lambda}_t) \in \arg \min_{\mathbf{x}_t \in \mathcal{X}} \mathcal{L}_t(\mathbf{x}_t, \hat{\lambda}_t). \quad (18b)$$

Notice that the multiplicative constant μ in (17b) allows for adaptation even in the steady state (i.e., $t \rightarrow \infty$), but the diminishing η_t is for online learning, as we shall discuss below.

For a better illustration of the effective price (17b), we call $\hat{\lambda}_t$ the statistically learnt price to obtain the exact optimal argument of the expected problem (11). We also call $\mu \mathbf{q}_t$ (which is exactly λ_t as shown in (13a)) the online adaptation term since it can track the instantaneous change of system statistics. Intuitively, a large μ will allow the effective policy to quickly respond to instantaneous variations so that the policy gains improved control of queue lengths, while a small μ puts more weight on statistical learning from historical samples so that the allocation strategy will incur less variance in the steady state. In this sense, the proposed LA-SDG can attain both statistical efficiency and adaptability.

Distinctly different from SDG with constant stepsize that involves a pure adaptation step, LA-SDG takes an additional learning step [cf. (13a)], which adopts gradient ascent with diminishing stepsize to find the “best empirical” dual variable from all observed network states. This pair of complementary gradient steps endows LA-SDG with its attractive properties. In its transient stage, the extra gradient evaluations and empirical dual variables accelerate the convergence speed of SDG; while in the steady stage, the empirical dual variable approaches the optimal multiplier, which significantly reduces the steady-state queue lengths.

Remark 1. Readers familiar with algorithms on statistical learning and stochastic network optimization can recognize their similarities and differences with LA-SDG.

(P1) SDG in [4] involves only the first part of LA-SDG (1st gradient), where the allocation policy purely relies on stochastic estimates of Lagrange multipliers or instantaneous queue lengths, i.e., $\gamma_t = \mu \mathbf{q}_t$. In contrast, LA-SDG further leverages statistical learning from streaming data.

(P2) Several schemes have been developed recently for statistical learning at scale to find $\hat{\lambda}_t$, namely, SAG in [25] and SAGA in [26]. However, directly applying $\gamma_t = \hat{\lambda}_t$ to allocate resources causes infeasibility. For a finite time t , $\hat{\lambda}_t$ is δ -optimal³ for (11), and the primal variable $\mathbf{x}_t(\hat{\lambda}_t)$ in turn is δ -feasible with respect to (7b) that is necessary for (3c). Since \mathbf{q}_t essentially accumulates online constraint violations of (7b), it will grow linearly with t and eventually become unbounded.

B. LA-SDG as a modified heavy-ball iteration

The heavy-ball iteration belongs to the family of momentum-based first-order methods, and has well-documented acceleration merits in the deterministic setting [27]. Motivated by its convergence speed in solving deterministic problems, stochastic heavy-ball methods have been also pursued recently [10], [13].

The stochastic version of the heavy-ball iteration is [13]

$$\lambda_{t+1} = \lambda_t + \mu \nabla \mathcal{D}_t(\lambda_t) + \beta(\lambda_t - \lambda_{t-1}), \forall t \quad (19)$$

³Iterate $\hat{\lambda}_t$ is δ -optimal if $\|\hat{\lambda}_t - \lambda^*\| \leq \mathcal{O}(\delta)$, and likewise for δ -feasibility.

where $\mu > 0$ is an appropriate constant stepsize, $\beta \in [0, 1)$ denotes the momentum factor, and the stochastic gradient $\nabla \mathcal{D}_t(\lambda_t)$ can be found by solving (13b) using heavy-ball iterate λ_t . This iteration exhibits attractive convergence rate during the initial stage, but its performance degrades in the steady state. Recently, the performance of momentum iterations (heavy-ball or Nesterov) with constant stepsize μ and momentum factor β , has been proved equivalent to SDG with constant $\mu/(1-\beta)$ per iteration [13]. Since SDG with a large stepsize converges fast at the price of considerable loss in optimality, the momentum methods naturally inherit these attributes.

To see the influence of the momentum term, consider expanding the iteration (19) as

$$\begin{aligned}\lambda_{t+1} &= \lambda_t + \mu \nabla \mathcal{D}_t(\lambda_t) + \beta(\lambda_t - \lambda_{t-1}) \\ &= \lambda_t + \mu \nabla \mathcal{D}_t(\lambda_t) + \beta [\mu \nabla \mathcal{D}_{t-1}(\lambda_{t-1}) + \beta(\lambda_{t-1} - \lambda_{t-2})] \\ &= \lambda_t + \underbrace{\mu \sum_{\tau=1}^t \beta^{t-\tau} \nabla \mathcal{D}_\tau(\lambda_\tau)}_{\text{accumulated gradient}} + \underbrace{\beta^t (\lambda_1 - \lambda_0)}_{\text{initial state}}.\end{aligned}\quad (20)$$

The stochastic heavy-ball method will accelerate convergence in the initial stage thanks to the accumulated gradients, and it will gradually forget the initial state. As t increases however, the algorithm also incurs a worst-case oscillation $\mathcal{O}(\mu/(1-\beta))$, which degrades performance in terms of objective values when compared to SDG with stepsize μ . This is in agreement with the theoretical analysis in [13, Theorem 11].

Different from standard momentum methods, LA-SDG nicely inherits the fast convergence in the initial stage, while reducing the oscillation of stochastic momentum methods in the steady state. To see this, consider two consecutive iterations (17b)

$$\gamma_{t+1} = \hat{\lambda}_{t+1} + \mu \mathbf{q}_{t+1} - \theta \quad (21a)$$

$$\gamma_t = \hat{\lambda}_t + \mu \mathbf{q}_t - \theta \quad (21b)$$

and subtract them, to arrive at

$$\begin{aligned}\gamma_{t+1} &= \gamma_t + \mu (\mathbf{q}_{t+1} - \mathbf{q}_t) + (\hat{\lambda}_{t+1} - \hat{\lambda}_t) \\ &= \gamma_t + \mu \nabla \mathcal{D}_t(\gamma_t) + (\hat{\lambda}_{t+1} - \hat{\lambda}_t), \quad \forall t.\end{aligned}\quad (22)$$

Here the equalities in (22) follows from $\nabla \mathcal{D}_t(\gamma_t) = \mathbf{A} \mathbf{x}_t(\gamma_t) + \mathbf{c}_t$ in \mathbf{q}_t recursion (16), and with a sufficiently large θ , the projection in (16) rarely (with sufficiently low probability) takes effect since the steady-state \mathbf{q}_t will hover around θ/μ ; see the details of Theorem 2 and the proof thereof.

Comparing the LA-SDG iteration (22) with the stochastic heavy-ball iteration (19), both of them correct the iterates using the stochastic gradient $\nabla \mathcal{D}_t(\gamma_t)$ or $\nabla \mathcal{D}_t(\lambda_t)$. However, LA-SDG incorporates the variation of a learning sequence (also known as a reference sequence) $\{\hat{\lambda}_t\}$ into the recursion of the main iterate γ_t , other than heavy-ball's momentum term $\beta(\lambda_t - \lambda_{t-1})$. Since the variation of learning iterate $\hat{\lambda}_t$ eventually diminishes as t increases, keeping the learning sequence enables LA-SDG to enjoy accelerated convergence in the initial (transient) stage compared to SDG, while avoiding large oscillation in the steady state compared to the stochastic heavy-ball method. We formally remark this observation next.

Remark 2. LA-SDG offers a fresh approach to designing stochastic optimization algorithms in a dynamic environment. While directly applying the momentum-based iteration to a

stochastic setting may lead to unsatisfactory steady-state performance, it is promising to carefully design a reference sequence that exactly converges to the optimal argument. Therefore, algorithms with improved convergence (e.g., the second-order method in [12]) can also be incorporated as a reference sequence to further enhance the performance of LA-SDG.

C. Complexity and distributed implementation of LA-SDG

This section introduces a fully distributed implementation of LA-SDG by exploiting the problem structure of network resource allocation. For notational brevity, collect the variables representing outgoing links from node i in $\mathbf{x}_t^i := \{x_t^{ij}, \forall j \in \mathcal{N}_i\}$ with \mathcal{N}_i denoting the index set of outgoing neighbors of node i . Let also $\mathbf{s}_t^i := [\phi_t^i; c_t^i]$ denote the random state at node i . It will be shown that the learning and allocation decision per time slot t is processed locally per node i based on its local state \mathbf{s}_t^i .

To this end, rewrite the Lagrangian minimization for a general dual variable $\lambda \in \mathbb{R}_+^I$ at time t as [cf. (17a) and (18b)]

$$\min_{\mathbf{x}_t \in \mathcal{X}} \sum_{i \in \mathcal{I}} \Psi^i(\mathbf{x}_t^i; \phi_t^i) + \sum_{i \in \mathcal{I}} \lambda^i (\mathbf{A}_{(i,:)} \mathbf{x}_t + c_t^i) \quad (23)$$

where λ^i is the i -th entry of vector λ , and $\mathbf{A}_{(i,:)}$ denotes the i -th row of the node-incidence matrix \mathbf{A} . Clearly, $\mathbf{A}_{(i,:)}$ selects entries of \mathbf{x}_t associated with the in- and out-links of node i . Therefore, the subproblem at node i is

$$\min_{\mathbf{x}_t^i \in \mathcal{X}^i} \Psi^i(\mathbf{x}_t^i; \phi_t^i) + \sum_{j \in \mathcal{N}_i} (\lambda^j - \lambda^i) x_t^{ji} \quad (24)$$

where \mathcal{X}^i is the feasible set of primal variable \mathbf{x}_t^i . In the case of (3d), the feasible set \mathcal{X} can be written as a Cartesian product of sets $\{\mathcal{X}^i, \forall i\}$, so that the projection of \mathbf{x}_t to \mathcal{X} is equivalent to separate projections of \mathbf{x}_t^i onto \mathcal{X}^i . Note that $\{\lambda^j, \forall j \in \mathcal{N}_i\}$ will be available at node i by exchanging information with the neighbors per time t . Hence, given the effective multipliers γ_t^j (j -th entry of γ_t) from its outgoing neighbors in $j \in \mathcal{N}_i$, node i is able to form an allocation decision $\mathbf{x}_t^i(\gamma_t)$ by solving the convex programs (24) with $\lambda^j = \gamma_t^j$; see also (17a). Needless to mention, q_t^i can be locally updated via (16), that is

$$q_{t+1}^i = \left[q_t^i + \left(\sum_{j:i \in \mathcal{N}_j} x_t^{ji}(\gamma_t) - \sum_{j \in \mathcal{N}_i} x_t^{ij}(\gamma_t) + c_t^i \right) \right]^+ \quad (25)$$

where $\{x_t^{ji}(\gamma_t)\}$ are the local measurements of arrival (departure) workloads from (to) its neighbors.

Likewise, the tentative primal variable $\mathbf{x}_t^i(\hat{\lambda}_t)$ can be obtained at each node locally by solving (24) using the current sample \mathbf{s}_t^i again with $\lambda^i = \hat{\lambda}_t^i$. By sending $\mathbf{x}_t^i(\hat{\lambda}_t)$ to its outgoing neighbors, node i can update the empirical multiplier $\hat{\lambda}_{t+1}^i$ via

$$\hat{\lambda}_{t+1}^i = \left[\hat{\lambda}_t^i + \eta_t \left(\sum_{j:i \in \mathcal{N}_j} x_t^{ji}(\hat{\lambda}_t) - \sum_{j \in \mathcal{N}_i} x_t^{ij}(\hat{\lambda}_t) + c_t^i \right) \right]^+ \quad (26)$$

which, together with the local queue length q_{t+1}^i , also implies that the next γ_{t+1}^i can be obtained locally.

Compared with the classic SDG recursion (13a)-(13b), the distributed implementation of LA-SDG incurs only a factor of two increase in computational complexity. Next, we will further

analytically establish that it can improve the delay of SDG by an order of magnitude with the same order of optimality gap.

V. OPTIMALITY AND STABILITY OF LA-SDG

This section presents performance analysis of LA-SDG, which will rely on the following four assumptions.

Assumption 1. The state \mathbf{s}_t is bounded and i.i.d. over time t .

Assumption 2. $\Psi_t(\mathbf{x}_t)$ in (3a) is σ -strongly convex, and has L_p -Lipschitz continuous gradient. Also, $\Psi_t(\mathbf{x}_t)$ is non-decreasing w.r.t. all entries of \mathbf{x}_t over a convex set \mathcal{X} .

Assumption 3. There exists a stationary policy $\chi(\cdot)$ satisfying $\chi(\mathbf{s}_t) \in \mathcal{X}$ for all \mathbf{s}_t , and $\mathbb{E}[\mathbf{A}\chi(\mathbf{s}_t) + \mathbf{c}_t] \leq -\zeta$, where $\zeta > 0$ is a slack vector constant.

Assumption 4. For any time t , the magnitude of the constraint is bounded, that is, $\|\mathbf{A}\mathbf{x}_t + \mathbf{c}_t\| \leq M$, $\forall \mathbf{x}_t \in \mathcal{X}$.

Assumption 1 is typical in stochastic network resource allocation [14], [15], [28], and can be relaxed to an ergodic and stationary setting following [20], [29]. Assumption 2 ensures uniqueness of the optimal solution. Non-decreasing costs with increased resources are easily guaranteed with exponential and quadratic functions. In addition, Assumption 2 ensures that the dual function has favorable properties, which are important to ensure stability. Assumption 3 is Slater's condition, which guarantees the existence of a bounded optimal Lagrange multiplier [22], and is also necessary for queue stability [4]. Assumption 4 guarantees boundedness of the gradient of the instantaneous dual function, which is common in performance analysis of stochastic gradient-type algorithms [30].

Building upon the desirable properties of the primal problem, we next show that the corresponding dual function satisfies both smoothness and quadratic growth properties [31], [32], which will be critical to the subsequent analysis.

Lemma 1. Under Assumption 2, the dual function $\mathcal{D}_t(\lambda)$ in (11) is L_d -smooth, where $L_d = \rho(\mathbf{A}^\top \mathbf{A})/\sigma$, and $\rho(\mathbf{A}^\top \mathbf{A})$ denotes the spectral radius of $\mathbf{A}^\top \mathbf{A}$. In addition, if the dual variable λ lies in a compact set, there always exists a small constant ϵ such that $\mathcal{D}(\lambda)$ satisfies the following quadratic growth property

$$\mathcal{D}(\lambda^*) - \mathcal{D}(\lambda) \geq \frac{\epsilon}{2} \|\lambda^* - \lambda\|^2 \quad (27)$$

where λ^* is the optimal multiplier for the dual problem (11).

Proof. See Appendix A. \square

We start with the convergence of the empirical dual variables $\hat{\lambda}_t$. Note that the update of $\hat{\lambda}_t$ is a classical learning process from historical data, and it is not affected by future resource allocation strategies. Therefore, the theoretical result on SDG with diminishing stepsize is directly applicable [30, Sec. 2.2].

Lemma 2. Let $\hat{\lambda}_t$ denote the empirical dual variable in Algorithm 1, and λ^* the optimal argument for the dual problem (11). If the stepsize is chosen as $\eta_t = \frac{\alpha D}{M\sqrt{t}}$, $\forall t$, with a constant $\alpha > 0$, a sufficient large constant $D > 0$, and M as in Assumption 4, then it holds that

$$\mathbb{E}[\mathcal{D}(\lambda^*) - \mathcal{D}(\hat{\lambda}_t)] \leq \max\{\alpha, \alpha^{-1}\} \frac{DM}{\sqrt{t}} \quad (28)$$

where the expectation is over all the random states \mathbf{s}_t up to t .

Lemma 2 asserts that using a diminishing stepsize, the dual function value converges sub-linearly to the optimal value in expectation. In principle, D is the radius of the feasible set for the dual variable λ [30, Sec. 2.2]. However, as the optimal multiplier λ^* is bounded according to Assumption 3, one can always estimate a large enough D , and the estimation error will only affect the constant of the sub-optimality bound (28) through the scalar α . The sub-optimality bound in Lemma 2 holds in expectation, which averages over all possible sample paths $\{\mathbf{s}_1, \dots, \mathbf{s}_t\}$.

As a complement to Lemma 2, the almost sure convergence of the empirical dual variables is established next to characterize the performance of each individual sample path.

Theorem 1. For the sequence of empirical multipliers $\{\hat{\lambda}_t\}$ in Algorithm 1, if the stepsizes are chosen as $\eta_t = \frac{\alpha D}{M\sqrt{t}}$, $\forall t$, with constants α, M, D defined in Lemma 2, it holds that

$$\lim_{t \rightarrow \infty} \hat{\lambda}_t = \lambda^*, \quad \text{w.p.1} \quad (29)$$

where λ^* is the optimal dual variable for the expected dual function minimization (11).

Proof. The proof follows the steps in [21, Proposition 8.2.13], which is omitted here. \square

Building upon the asymptotic convergence of empirical dual variables for statistical learning, it becomes possible to analyze the online performance of LA-SDG. Clearly, the online resource allocation \mathbf{x}_t is a function of the effective dual variable γ_t and the instantaneous network state \mathbf{s}_t [cf. (17a)]. Therefore, the next step is to show that the effective dual variable γ_t also converges to the optimal argument of the expected problem (11), which would establish that the online resource allocation \mathbf{x}_t is asymptotically optimal. However, directly analyzing the trajectory of γ_t is nontrivial, because the queue length $\{\mathbf{q}_t\}$ is coupled with the reference sequence $\{\hat{\lambda}_t\}$ in γ_t . To address this issue, rewrite the recursion of γ_t as

$$\gamma_{t+1} = \gamma_t + (\hat{\lambda}_{t+1} - \hat{\lambda}_t) + \mu(\mathbf{q}_{t+1} - \mathbf{q}_t), \quad \forall t \quad (30)$$

where the update of γ_t depends on the variations of $\hat{\lambda}_t$ and \mathbf{q}_t . We will first study the asymptotic behavior of queue lengths \mathbf{q}_t , and then derive the analysis of γ_t using the convergence of $\hat{\lambda}_t$ in (29), and the recursion (30).

Define the time-varying target $\tilde{\theta}_t = \lambda^* - \hat{\lambda}_t + \theta$, which is the optimality residual of statistical learning $\lambda^* - \hat{\lambda}_t$ plus the bias-control variable θ . Per Theorem 1, it readily follows that $\lim_{t \rightarrow \infty} \tilde{\theta}_t = \theta$, w.p.1. By showing that \mathbf{q}_t is attracted towards the time-varying target $\tilde{\theta}_t/\mu$, we will further derive the stability of queue lengths.

Lemma 3. With \mathbf{q}_t and μ denoting queue length and stepsize, there exists a constant $B = \Theta(1/\sqrt{\mu})$, and a finite time $T_B < \infty$, such that for all $t \geq T_B$, if $\|\mathbf{q}_t - \tilde{\theta}_t/\mu\| > B$, it holds in LA-SDG that

$$\mathbb{E}[\|\mathbf{q}_{t+1} - \tilde{\theta}_t/\mu\| | \mathbf{q}_t] \leq \|\mathbf{q}_t - \tilde{\theta}_t/\mu\| - \sqrt{\mu}, \quad \text{w.p.1.} \quad (31)$$

Proof. See Appendix B. \square

Lemma 3 reveals that when \mathbf{q}_t is large and deviates from the time-varying target $\hat{\theta}_t/\mu$, it will be bounced back towards the target in the next time slot. Upon establishing this drift behavior of queues, we are on track to establish queue stability.

Theorem 2. *With $\mathbf{q}_t, \theta, \mu$ defined in (17b), there exists a constant $\tilde{B} = \Theta(1/\sqrt{\mu})$ such that the queue length under LA-SDG converges to a neighborhood of θ/μ as*

$$\liminf_{t \rightarrow \infty} \|\mathbf{q}_t - \theta/\mu\| \leq \tilde{B}, \text{ w.p.1.} \quad (32a)$$

In addition, if we choose $\theta = \mathcal{O}(\sqrt{\mu} \log^2(\mu))$, the long-term average expected queue length satisfies

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{q}_t] = \mathcal{O}\left(\frac{\log^2(\mu)}{\sqrt{\mu}}\right), \text{ w.p.1.} \quad (32b)$$

Proof. See Appendix C. \square

Theorem 2 in (32a) asserts that the sequence of queue iterates converges (in the infimum sense) to a neighborhood of θ/μ , where the radius of neighborhood region scales as $1/\sqrt{\mu}$. In addition to the sample path result, (32b) demonstrates that with a specific choice of θ , the queue length averaged over all sample paths will be $\mathcal{O}(\log^2(\mu)/\sqrt{\mu})$. Together with Theorem 1, it suffices to have the effective dual variable converge to a neighborhood of the optimal multiplier λ^* ; that is, $\liminf_{t \rightarrow \infty} \gamma_t = \lambda^* + \mu \mathbf{q}_t - \theta = \lambda^* + \mathcal{O}(\sqrt{\mu})$, w.p.1. Notice that the SDG iterate λ_t in (13a) will also converge to a neighborhood of λ^* . Therefore, intuitively LA-SDG will behave similar to SDG in the steady state, and its asymptotic performance follows from that of SDG. However, the difference is that through a careful choice of θ , for a sufficiently small μ , LA-SDG can improve the queue length $\mathcal{O}(1/\mu)$ under SDG by an order of magnitude.

In addition to feasibility, we formally establish in the next theorem that LA-SDG is asymptotically near-optimal.

Theorem 3. *Let Ψ^* be the optimal objective value of (3) under any feasible policy with distribution information about the state fully available. If the control variable is chosen as $\theta = \mathcal{O}(\sqrt{\mu} \log^2(\mu))$, then with a sufficiently small μ , the proposed LA-SDG algorithm yields a near-optimal solution for (3) in the sense that*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t(\gamma_t))] \leq \Psi^* + \mathcal{O}(\mu), \text{ w.p.1} \quad (33)$$

where $\mathbf{x}_t(\gamma_t)$ denotes the real-time operations obtained from the Lagrangian minimization (17a).

Proof. See Appendix D. \square

Combining Theorems 2 and 3, we are ready to state that by setting $\theta = \mathcal{O}(\sqrt{\mu} \log^2(\mu))$, LA-SDG is asymptotically $\mathcal{O}(\mu)$ -optimal with an average queue length $\mathcal{O}(\log^2(\mu)/\sqrt{\mu})$. This result implies that LA-SDG is able to achieve a near-optimal cost-delay tradeoff $[\mu, \log^2(\mu)/\sqrt{\mu}]$; see [4], [19]. Comparing with the standard tradeoff $[\mu, 1/\mu]$ under SDG, the learn-and-adapt design of LA-SDG markedly improves the online performance in terms of delay. Note that a better tradeoff $[\mu, \log^2(\mu)]$ has been derived in [15] under the so-termed local polyhedral

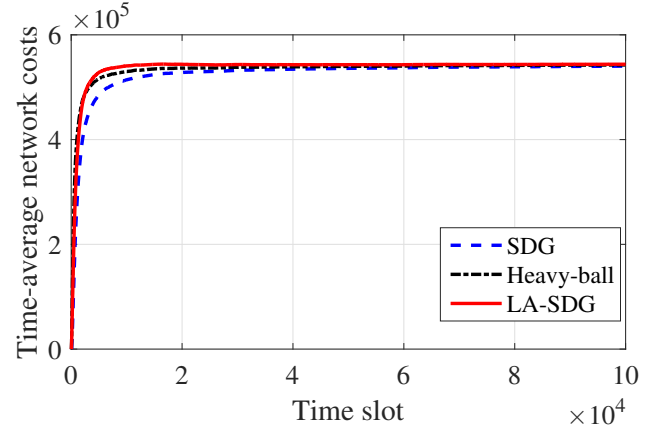


Fig. 2. Comparison of time-averaged network costs ($\beta = 0.5, \mu = 0.2$).

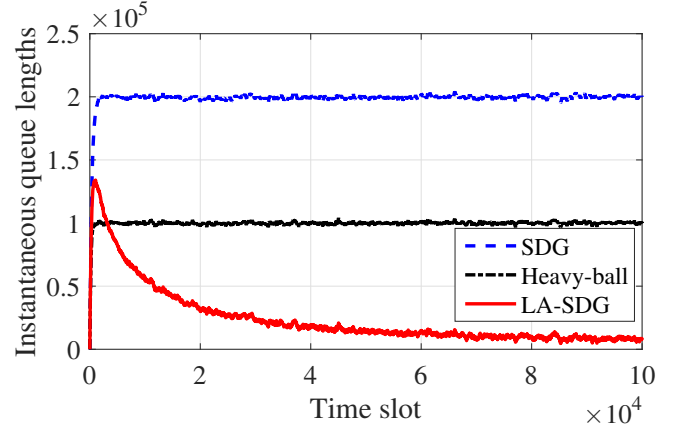


Fig. 3. Queue lengths summed over all nodes ($\beta = 0.5, \mu = 0.2$).

assumption. Observe though, that the considered setting in [15] is different from the one here. While the network state set \mathcal{S} and the action set \mathcal{X} in [15] are discrete and countable, LA-SDG allows continuous \mathcal{S} and \mathcal{X} with possibly infinite elements, and still be amenable to efficient and scalable online operations.

VI. NUMERICAL TESTS

This section presents numerical tests to confirm the analytical claims and demonstrate the merits of the proposed approach. We consider the geographical load balancing network of Section II-B with $K = 10$ data centers, and $J = 10$ mapping nodes. Performance is tested in terms of the time-averaged instantaneous network cost in (4), namely

$$\Psi_t(\mathbf{x}_t) := \sum_{k \in \mathcal{K}} p_t^k ((y_t^k)^2 - c_t^k) + \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} b_t^{jk} (x_t^{jk})^2 \quad (34)$$

where the energy price p_t^k is uniformly distributed over $[10, 30]$; samples of the renewable supply $\{e_t^k\}$ are generated uniformly over $[10, 100]$; and the per-unit bandwidth cost is set to $b_t^{jk} = 40/\bar{x}^{jk}, \forall k, j$, with bandwidth limits $\{\bar{x}^{jk}\}$ generated from a uniform distribution within $[100, 200]$. The capacities at data centers $\{\bar{y}^k\}$ are uniformly generated from $[100, 200]$. The delay-tolerant workloads $\{c_t^j\}$ arrive at each mapping node j according to a uniform distribution over $[10, 100]$. Clearly, the cost (34) and the state \mathbf{s}_t here satisfy Assumptions 1 and 2. Finally, the stepsize is $\eta_t = 1/\sqrt{t}, \forall t$, the trade-off variable is $\mu = 0.2$, and the bias correction vector is chosen

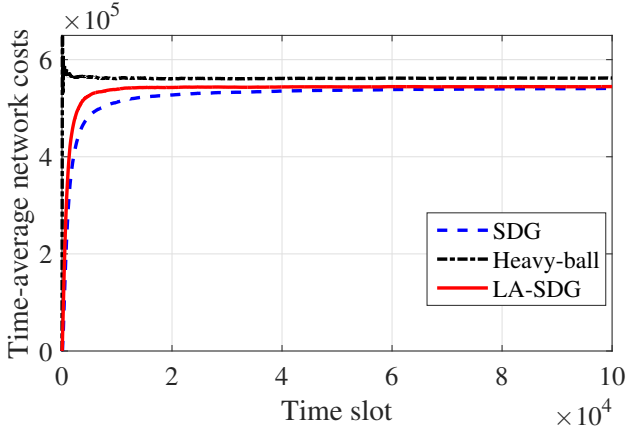


Fig. 4. Comparison of time-averaged network costs ($\beta = 0.99, \mu = 0.2$).

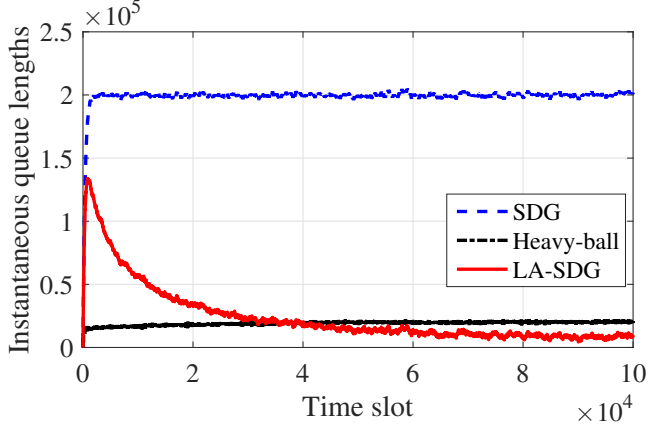


Fig. 5. Queue lengths summed over all nodes ($\beta = 0.99, \mu = 0.2$).

as $\theta = 50\sqrt{\mu} \log^2(\mu)\mathbf{1}$ by default, but manually tuned in Figs. 7-8. We introduce two benchmarks: SDG in (13a) (see e.g., [4]), and the projected stochastic heavy-ball in (19) and $\beta = 0.5$ by default (see e.g., [10]).

Performance is compared in terms of the time-averaged cost, and the instantaneous queue length in Figs. 2 and 3. For the network cost, the three algorithms converge to almost the same value. LA-SDG and heavy-ball exhibit faster convergence than SDG as their running-average costs quickly arrive at the optimal operating phase by leveraging the learning process or the momentum acceleration. Under this parameter configuration ($\beta = 0.5, \mu = 0.2$), LA-SDG exhibits a much lower delay as its aggregated queue length is only 10% of that for heavy-ball and 4% of that for SDG. Clearly, our learn-and-adapt procedure improves the delay performance.

The same metrics are also compared by setting $\beta = 0.99, \mu = 0.2$; see Figs. 4 and 5. With a larger momentum factor $\beta = 0.99$, the heavy-ball method exhibits a pronounced optimality loss relative to SDG in Fig. 4, while LA-SDG still converges to a time-average cost similar to that of SDG. However, by using a larger β , the heavy-ball method incurs a much lower queue length relative to that in Fig. 3. Comparing to the LA-SDG's queue length however, it is still much larger in the steady state.

Recall that the instantaneous resource allocation can be viewed as a function of the dual variable; see Proposition 1. Hence, the performance differences in Figs. 2-3 and 4-5 can

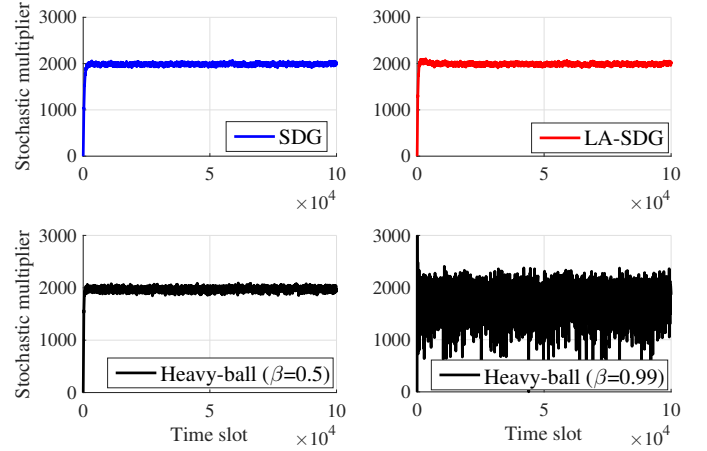


Fig. 6. The evolution of stochastic multipliers at mapping node 1 ($\mu = 0.2$).

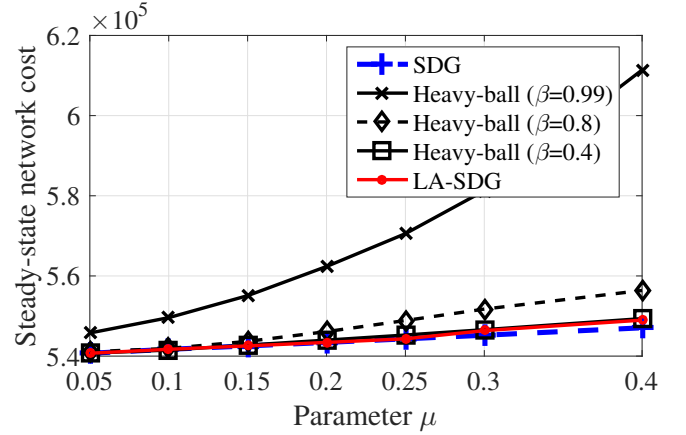


Fig. 7. Comparison of steady-state network costs (after 10^6 slots).

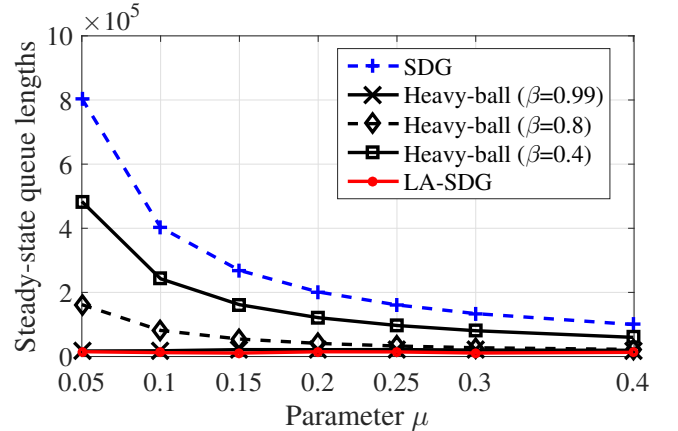


Fig. 8. Steady-state queue lengths summed over all nodes (after 10^6 slots).

be also anticipated by the different behavior of dual variables. In Fig. 6, the evolution of stochastic dual variables are plotted under two considered settings; that is the dual iterate in (13a) for SDG, the momentum iteration in (19) for the heavy-ball method, and the effective multiplier in (17b) for LA-SDG. As illustrated in (20), the performance of momentum iterations is similar to SDG with larger stepsize $\mu/(1 - \beta)$. This is corroborated by Fig. 6, where the stochastic momentum iterate with $\beta = 0.5$ behaves similar to the dual iterates of SDG and LA-SDG, but its oscillation becomes prohibitively high with a larger factor $\beta = 0.99$, which nicely explains the higher cost in Fig. 4.

Since the cost-delay performance is sensitive to the choice of parameters μ and β , extensive experiments are further conducted among three algorithms using different values of μ and β in Figs. 7 and 8. The steady-state performance is evaluated by running algorithms for sufficiently long time, up to 10^6 slots. The steady-state costs of all three algorithms increase as μ becomes larger, and the costs of LA-SDG and the heavy-ball with small momentum factor $\beta = 0.4$ are close to that of SDG, while the costs of the heavy-ball with larger momentum factors $\beta = 0.8$ and $\beta = 0.99$ are much larger than that of SDG. Considering steady-state queue lengths (network delay), LA-SDG exhibits an order of magnitude lower amount than those of SDG and the heavy-ball with small β , under all choices of μ . Note that the heavy-ball with a sufficiently large factor $\beta = 0.99$ also has a very low queue length, but it incurs a much higher cost than LA-SDG in Fig. 7 due to a higher steady-state oscillation observed in Fig. 6.

VII. CONCLUDING REMARKS

Fast convergent resource allocation and low service delay are highly desirable attributes of stochastic network management approaches. Leveraging recent advances in online learning and momentum-based optimization, a novel online approach termed LA-SDG was developed in this paper. LA-SDG learns the network state statistics through an additional sample recourse procedure. The associated novel iteration can be nicely interpreted as a modified heavy-ball recursion with an extra correction step to mitigate steady-state oscillations. It was analytically established that LA-SDG achieves a near-optimal cost-delay tradeoff $[\mu, \log^2(\mu)/\sqrt{\mu}]$, which is better than $[\mu, 1/\mu]$ of SDG, at the cost of only one extra gradient evaluation per new datum. Our future research agenda includes novel approaches to further hedge against non-stationarity, and improved learning schemes to uncover other valuable statistical patterns from historical data.

ACKNOWLEDGEMENT

The authors would like to thank Profs. Xin Wang, Longbo Huang and Jia Liu for helpful discussions.

APPENDIX

Let us first state a simple but useful property regarding the primal-dual problems (7) and (11).

Proposition 3. *Under Assumptions 1-3, for the constrained optimization (7) with the optimal policy $\chi^*(\cdot)$ and its optimal Lagrange multiplier λ^* , it holds that $\mathbb{E}[\mathbf{A}\mathbf{x}_t^* + \mathbf{c}_t] = \mathbf{0}$ with $\mathbf{x}_t^* = \chi^*(\mathbf{s}_t)$, and accordingly that $\nabla \mathcal{D}(\lambda^*) = \mathbf{0}$.*

Proof. With λ^* denoting the optimal Lagrange multiplier with (7b), the Karush-Kuhn-Tucker (KKT) conditions [22] are

$$(\mathbb{E}[\nabla \Psi_t(\mathbf{x}_t^*)] + \mathbf{A}^\top \lambda^*)^\top (\mathbb{E}[\mathbf{x}_t - \mathbf{x}_t^*]) \geq 0, \forall \mathbf{x}_t \in \mathcal{X} \quad (35a)$$

$$(\lambda^*)^\top \mathbb{E}[\mathbf{A}\mathbf{x}_t^* + \mathbf{c}_t] = 0 \quad (35b)$$

$$\mathbb{E}[\mathbf{A}\mathbf{x}_t^* + \mathbf{c}_t] \leq \mathbf{0}; \lambda^* \geq \mathbf{0} \quad (35c)$$

where (35a) is the optimality condition of Lagrangian minimization, (35b) is the complementary slackness condition, and (35c) are the primal and dual feasibility conditions.

To establish the claim, let us first assume that there exists entry k that the inequality constraint (7b) is not active; i.e., $\mathbb{E}[\mathbf{A}_{(k,:)}\mathbf{x}_t^* + c_t^k] = -\zeta$ with the constant $\zeta > 0$, and $\mathbf{A}_{(k,:)}$ denoting the k -th row of \mathbf{A} . As each row of \mathbf{A} has at least one entry equal to -1 , we collect all indices of entries at k -th row with value -1 in set \mathcal{E}_k^{-1} so that $\mathbf{A}_{(k,e)} = -1, \forall e \in \mathcal{E}_k^{-1}$.

Since \mathbf{x}_t^* is feasible, we have $\mathbf{x}_t^* \geq \mathbf{0}$, and thus

$$\mathbb{E}[\mathbf{A}_{(k,:)}\mathbf{x}_t^* + c_t^k] = \mathbb{E}\left[\sum_{e \in \mathcal{E}_k^{-1}} \mathbf{A}_{(k,e)}(x_t^e)^* + c_t^k\right] = -\zeta \quad (36)$$

which implies that

$$\mathbb{E}\left[\sum_{e \in \mathcal{E}_k^{-1}} (x_t^e)^*\right] = \zeta + \mathbb{E}[c_t^k + \sum_{e \in \mathcal{E}_k \setminus \mathcal{E}_k^{-1}} \mathbf{A}_{(k,e)}(x_t^e)^*] > 0. \quad (37)$$

According (35b), it further follows that $(\lambda^k)^* = 0$ since $(\lambda^k)^* \cdot \mathbb{E}[\mathbf{A}_{(k,:)}\mathbf{x}_t^* + c_t^k] = -(\lambda^k)^* \cdot \zeta = 0$. Now we are on track to show that it contradicts with (35a). Since $\mathbb{E}[\sum_{e \in \mathcal{E}_k^{-1}} (x_t^e)^*] > 0$, there exists at least an index j such that $\mathbb{E}[(x_t^j)^*] > 0, j \in \mathcal{E}_k^{-1}$. Choose $\mathbb{E}[\mathbf{x}_t]$ with $\mathbb{E}[x_t^j] = \mathbb{E}[(x_t^j)^*], \forall j \neq j$ and $\mathbb{E}[x_t^j] = 0$, to have $\mathbb{E}[\mathbf{x}_t - \mathbf{x}_t^*] = [0, \dots, -\mathbb{E}[(x_t^j)^*], \dots, 0]^\top$. Hence, we arrive at (with $\nabla_j \Psi_t(\mathbf{x}_t^*)$ denoting j -th entry of gradient)

$$\begin{aligned} & (\mathbb{E}[\nabla \Psi_t(\mathbf{x}_t^*)] + \mathbf{A}^\top \lambda^*)^\top \mathbb{E}[\mathbf{x}_t - \mathbf{x}_t^*] \\ &= -\mathbb{E}\left[\nabla_j \Psi_t(\mathbf{x}_t^*)(x_t^j)^* - \sum_{i \in \mathcal{I}} (\lambda^i)^* \mathbf{A}_{(i,j)}(x_t^j)^*\right] \\ &\stackrel{(a)}{=} \underbrace{-\mathbb{E}[\nabla_j \Psi_t(\mathbf{x}_t^*)(x_t^j)^*]}_{< 0} - \sum_{i \in \mathcal{I} \setminus k} \underbrace{(\lambda^i)^* \mathbf{A}_{(i,j)} \mathbb{E}[(x_t^j)^*]}_{\geq 0} < 0 \quad (38) \end{aligned}$$

where (a) uses $(\lambda^k)^* = 0$; the first bracket follows from Assumption 2 since $\nabla_j \Psi_t(\mathbf{x}_t^*)$ is monotonically increasing and $\nabla_j \Psi_t(\mathbf{x}_t^*) \geq 0$, thus for $\mathbb{E}[(x_t^j)^*] > 0$ it follows $\mathbb{E}[\nabla_j \Psi_t(\mathbf{x}_t^*)] > 0$; and the second bracket follows that $\lambda^* \geq \mathbf{0}$ and each column of \mathbf{A} has at most one -1 and $\mathbf{A}_{(k,j)} = -1$. The proof is then complete since (38) contradicts (35a). \square

A. Proof of Lemma 1

Proof of Lipschitz continuity: Under Assumption 2, the primal objective $\Psi_t(\mathbf{x}_t)$ is σ -strongly convex, and the smooth constant of the dual function $\mathcal{D}_t(\lambda)$, or equivalently, the Lipschitz constant of gradient $\nabla \mathcal{D}_t(\lambda)$ directly follows from [9, Lemma II.2], which equals to $L_d = \rho(\mathbf{A}^\top \mathbf{A})/\sigma$, with $\rho(\mathbf{A}^\top \mathbf{A})$ denoting the maximum eigenvalue of $\mathbf{A}^\top \mathbf{A}$. We omit the derivations of this result, and refer readers to that in [9].

Supporting lemmas for quadratic growth: To prove the quadratic growth property (27), we introduce an error bound, which describes the local property of the dual function $\mathcal{D}(\lambda)$.

Lemma 4. [31, Lemma 2.3] *Consider the dual function in (10) and the feasible set \mathcal{X} in (3d) with only linear constraints. For any λ satisfying $\mathcal{D}(\lambda) > -\infty$ and $\|\nabla \mathcal{D}(\lambda)\| \leq \delta$, we have*

$$\|\lambda^* - \lambda\| \leq \xi \|\nabla \mathcal{D}(\lambda)\| \quad (39)$$

where the scalar ξ depends on the matrix \mathbf{A} as well the constants σ, L_p and L_d introduced in Assumption 2.

Lemma 4 states a *local* error bound for the dual function $\mathcal{D}(\lambda)$. The error bound is “local” since it holds only for λ close

enough to the optimum λ^* , i.e., $\|\nabla \mathcal{D}(\lambda)\| \leq \delta$. Following the arguments in [31] however, if the dual iterate λ is artificially confined to a compact set Λ such that $\|\lambda\| \leq D$ with D denoting the radius of Λ ,⁴ then for the case $\|\nabla \mathcal{D}(\lambda)\| \geq \delta$, the ratio $\|\lambda^* - \lambda\|/\|\nabla \mathcal{D}(\lambda)\| \leq D/\delta$, which implies the existence of ξ satisfying (39) for any $\lambda \in \Lambda$. Lemma 4 is important for establishing linear convergence rate without strong convexity [31], [33]. Remarkably, we will show next that this error bound is also critical to characterize the steady-state behavior of our LA-SDG scheme.

Building upon Lemma 4, we next show that the ensemble dual function $\mathcal{D}(\lambda)$ also satisfies the so-termed Polyak-Lojasiewicz (PL) condition [32].

Lemma 5. *Under Assumption 2, the local error-bound in (39) implies the following PL condition, namely*

$$\mathcal{D}(\lambda^*) - \mathcal{D}(\lambda) \leq \frac{L_d \xi^2}{2} \|\nabla \mathcal{D}(\lambda)\|^2 \quad (40)$$

where L_d is the Lipschitz constant of the dual gradient and ξ is as in (39).

Proof. Using the L_d -smoothness of the dual function $\mathcal{D}(\lambda)$, we have for any λ and $\varphi \in \mathbb{R}_+^I$ that

$$\mathcal{D}(\varphi) \leq \mathcal{D}(\lambda) - \langle \nabla \mathcal{D}(\varphi), \lambda - \varphi \rangle + \frac{L_d}{2} \|\lambda - \varphi\|^2. \quad (41)$$

Choosing $\varphi = \lambda^*$, and using Proposition 3 such that $\nabla \mathcal{D}(\lambda^*) = \mathbf{0}$, we have

$$\mathcal{D}(\lambda^*) \leq \mathcal{D}(\lambda) + \frac{L_d}{2} \|\lambda - \lambda^*\|^2 \stackrel{(a)}{\leq} \mathcal{D}(\lambda) + \frac{L_d \xi^2}{2} \|\nabla \mathcal{D}(\lambda)\|^2 \quad (42)$$

where inequality (a) uses the local error-bound in (39). \square

Proof of quadratic growth: The proof follows the main steps of that in [32]. Building upon Lemma 5, we next prove Lemma 1. Define a function of the dual variable λ as $g(\lambda) := \sqrt{\mathcal{D}(\lambda^*) - \mathcal{D}(\lambda)}$. With the PL condition in (40), and Λ^* denoting the set of optimal multipliers for (11), we have for any $\lambda \notin \Lambda^*$ that

$$\|\nabla g(\lambda)\|^2 = \frac{\|\nabla \mathcal{D}(\lambda)\|^2}{\mathcal{D}(\lambda^*) - \mathcal{D}(\lambda)} \geq \frac{2}{L_d \xi^2} \quad (43)$$

which implies that $\|\nabla g(\lambda)\| \geq \sqrt{2/(L_d \xi^2)}$.

For any $\lambda_0 \notin \Lambda^*$, consider the following differential equation⁵

$$\begin{cases} \frac{d\lambda(\tau)}{d\tau} = -\nabla g(\lambda(\tau)) & (44a) \\ \lambda(\tau = 0) = \lambda_0 & (44b) \end{cases}$$

which describes the continuous trajectory of $\{\lambda(\tau)\}$ starting from λ_0 along the direction of $-\nabla g(\lambda(\tau))$. By using $\|\nabla g(\lambda)\| \geq \sqrt{2/(L_d \xi^2)}$, it follows that $\nabla g(\lambda)$ is bounded below; thus, the differential equation (44) guarantees that we sufficiently reduce the value of function $g(\lambda)$, and $\lambda(\tau)$ will eventually reach Λ^* .

⁴Since the optimal multiplier is bounded per Assumption 3, one can safely find a large set Λ with radius D to project dual iterates during optimization.

⁵The time index in the proof of Lemma 1 is not related to the online optimization process, but it is useful to find the structure of the dual function.

In other words, there exists a time T such that $\lambda(T) \in \Lambda^*$. Formally, for $\tau > T$, we have

$$\begin{aligned} g(\lambda_0) - g(\lambda_\tau) &= \int_{\lambda_\tau}^{\lambda_0} \langle \nabla g(\lambda), d\lambda \rangle \\ &= - \int_{\lambda_0}^{\lambda_\tau} \langle \nabla g(\lambda), d\lambda \rangle = - \int_0^T \left\langle \nabla g(\lambda), \frac{d\lambda(\tau)}{d\tau} \right\rangle d\tau \\ &= \int_0^T \|\nabla g(\lambda(\tau))\|^2 d\tau \geq \int_0^T \frac{2}{L_d \xi^2} d\tau = \frac{2T}{L_d \xi^2}. \end{aligned} \quad (45)$$

Since $g(\lambda) \geq 0$, $\forall \lambda$, we have $T \leq g(\lambda_0) L_d \xi^2 / 2$, which implies that there exists a finite time T such that $\lambda_\tau \in \Lambda^*$. On the other hand, the path length of trajectory $\{\lambda(\tau)\}$ will be longer than the projection distance between λ_0 and the closest point in Λ^* denoted as λ^* , that is,

$$\int_0^T \left\| \frac{d\lambda(\tau)}{d\tau} \right\| d\tau = \int_0^T \|\nabla g(\lambda(\tau))\| d\tau \geq \|\lambda_0 - \lambda^*\| \quad (46)$$

and thus we have from (45) that

$$\begin{aligned} g(\lambda_0) - g(\lambda_\tau) &= \int_0^T \|\nabla g(\lambda(\tau))\|^2 d\tau \\ &\geq \int_0^T \|\nabla g(\lambda(\tau))\| \sqrt{\frac{2}{L_d \xi^2}} d\tau \stackrel{(b)}{\geq} \sqrt{\frac{2}{L_d \xi^2}} \|\lambda_0 - \lambda^*\| \end{aligned} \quad (47)$$

where (b) follows from (46). Choosing T such that $g(\lambda_T) = 0$, we have

$$g(\lambda_0) \geq \sqrt{\frac{2}{L_d \xi^2}} \|\lambda_0 - \lambda^*\|. \quad (48)$$

Squaring both sides of (48), the proof is complete, since ϵ is defined as $\epsilon := 2/(L_d \xi^2)$ and λ_0 can be any point outside the set of optimal multipliers.

B. Proof of Lemma 3

Since $\hat{\lambda}_t$ converges to λ^* , w.p.1 according to Theorem 1, there exists a finite time T_θ such that for $t > T_\theta$, we have $\|\lambda^* - \hat{\lambda}_t\| \leq \|\theta\|$. In such case, it follows that $\tilde{\theta}_t = \lambda^* - \hat{\lambda}_t + \theta \geq \mathbf{0}$, since $\theta \geq \mathbf{0}$. Therefore, we have

$$\begin{aligned} \|\mathbf{q}_{t+1} - \tilde{\theta}_t/\mu\|^2 &= \|[\mathbf{q}_t + \mathbf{A}\mathbf{x}_t + \mathbf{c}_t]^+ - [\tilde{\theta}_t/\mu]^+\|^2 \\ &\stackrel{(a)}{\leq} \|\mathbf{q}_t + \mathbf{A}\mathbf{x}_t + \mathbf{c}_t - \tilde{\theta}_t/\mu\|^2 \\ &\stackrel{(b)}{\leq} \|\mathbf{q}_t - \tilde{\theta}_t/\mu\|^2 + 2(\mathbf{q}_t - \tilde{\theta}_t/\mu)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2 \end{aligned} \quad (49)$$

where (a) comes from the non-expansive property of the projection operator, and (b) is due to the upper bound M in Assumption 4.

The RHS of (49) can be upper bounded by

$$\begin{aligned} &\|\mathbf{q}_t - \tilde{\theta}_t/\mu\|^2 + 2(\mathbf{q}_t - \tilde{\theta}_t/\mu)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2 \\ &\stackrel{(c)}{=} \|\mathbf{q}_t - \tilde{\theta}_t/\mu\|^2 + \frac{2}{\mu} (\gamma_t - \lambda^*)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2 \end{aligned} \quad (50)$$

where (c) uses the definitions $\tilde{\theta}_t := \lambda^* - \hat{\lambda}_t + \theta$, and $\gamma_t := \hat{\lambda}_t + \mu\mathbf{q}_t - \theta$. Since $\mathbf{A}\mathbf{x}_t + \mathbf{c}_t$ is the stochastic subgradient of the concave function $\mathcal{D}(\lambda)$ at $\lambda = \gamma_t$ [cf. (17a)], we have

$$\mathbb{E} \left[(\gamma_t - \lambda^*)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] \leq \mathcal{D}(\gamma_t) - \mathcal{D}(\lambda^*). \quad (51)$$

Taking expectations on (49)-(50) over the random state s_t conditioned on \mathbf{q}_t and using (51), we arrive at

$$\mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 \right] \leq \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 + \frac{2}{\mu} (\mathcal{D}(\gamma_t) - \mathcal{D}(\boldsymbol{\lambda}^*)) + M^2 \quad (52)$$

where we use the fact that $\mathcal{D}(\boldsymbol{\lambda}) := \mathbb{E}[\mathcal{D}_t(\boldsymbol{\lambda})]$ in (11). Using the quadratic growth property of $\mathcal{D}(\boldsymbol{\lambda})$ in (27) of Lemma 1, the recursion (52) further leads to

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 \right] &\leq \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 - \frac{2\epsilon}{\mu} \|\gamma_t - \boldsymbol{\lambda}^*\|^2 + M^2 \\ &\stackrel{(d)}{\leq} \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 - 2\mu\epsilon \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 + M^2 \end{aligned} \quad (53)$$

where equality (d) uses the definitions $\tilde{\boldsymbol{\theta}}_t := \boldsymbol{\lambda}^* - \hat{\boldsymbol{\lambda}}_t + \boldsymbol{\theta}$ and $\gamma_t := \hat{\boldsymbol{\lambda}}_t + \mu\mathbf{q}_t - \boldsymbol{\theta}$, implying that $\gamma_t - \boldsymbol{\lambda}^* = \mu\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t$.

Now considering (cf. (53))

$$-2\mu\epsilon \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 + M^2 \leq -2\sqrt{\mu} \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| + \mu \quad (54)$$

and plugging it back into (53) yields

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 \right] &\leq \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 - 2\sqrt{\mu} \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| + \mu \\ &= \left(\|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| - \sqrt{\mu} \right)^2. \end{aligned} \quad (55)$$

By the convexity of $(\cdot)^2$, we further arrive at

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\| \right]^2 &\leq \mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\|^2 \right] \\ &\leq \left(\|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| - \sqrt{\mu} \right)^2 \end{aligned} \quad (56)$$

which directly implies the argument (31) in the lemma. By checking Vieta's formulas for second-order equations, there exists $B = \Theta(\frac{1}{\sqrt{\mu}})$ such that for $\|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| > B$, inequality (54) holds, and thus the lemma follows readily.

C. Proof of Theorem 2

Proof of (32a) in Theorem 2: Theorem 1 asserts that $\hat{\boldsymbol{\lambda}}_t$ eventually converges to the optimum $\boldsymbol{\lambda}^*$, w.p.1. Hence, there always exists a finite time T_ρ and an arbitrarily small ρ such that for $t > T_\rho$, it holds that $\|\boldsymbol{\lambda}^*/\mu - \hat{\boldsymbol{\lambda}}_t/\mu\| \leq \rho$. Using the definition $\tilde{\boldsymbol{\theta}}_t = \boldsymbol{\lambda}^* - \hat{\boldsymbol{\lambda}}_t + \boldsymbol{\theta}$, it then follows by the triangle inequality that

$$\left| \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| - \|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| \right| \leq \|\boldsymbol{\lambda}^*/\mu - \hat{\boldsymbol{\lambda}}_t/\mu\| \leq \rho \quad (57)$$

which also holds for \mathbf{q}_{t+1} .

Using (57) and the conditional drift (31) in Lemma 3, for $t > T_\rho$ and $\|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| > B = \Theta(1/\sqrt{\mu})$, it holds that

$$\begin{aligned} \mathbb{E} \left[\|\mathbf{q}_{t+1} - \boldsymbol{\theta}/\mu\| \mid \mathbf{q}_t \right] &\leq \mathbb{E} \left[\|\mathbf{q}_{t+1} - \tilde{\boldsymbol{\theta}}_t/\mu\| \mid \mathbf{q}_t \right] + \rho \\ &\leq \left| \|\mathbf{q}_t - \tilde{\boldsymbol{\theta}}_t/\mu\| - \sqrt{\mu} \right| + \rho \leq \|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| - \sqrt{\mu} + 2\rho. \end{aligned} \quad (58)$$

Choosing ρ such that $\sqrt{\mu} := \sqrt{\mu} - 2\rho < 0$, then for $t > T_\rho$ and $\|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| > \tilde{B} := B + \rho = \Theta(1/\sqrt{\mu})$, we have

$$\mathbb{E} \left[\|\mathbf{q}_{t+1} - \boldsymbol{\theta}/\mu\| \mid \mathbf{q}_t \right] \leq \|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| - \sqrt{\mu}. \quad (59)$$

Leveraging (59), we first show (32a) by constructing a super-martingale. Define the stochastic process a_t as

$$a_t := \|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| \cdot \mathbb{1} \left\{ \min_{\tau \leq t} \|\mathbf{q}_\tau - \boldsymbol{\theta}/\mu\| > \tilde{B} \right\}, \quad \forall t \quad (60a)$$

and likewise the stochastic process b_t as

$$b_t := \sqrt{\mu} \cdot \mathbb{1} \left\{ \min_{\tau \leq t} \|\mathbf{q}_\tau - \boldsymbol{\theta}/\mu\| > \tilde{B} \right\}, \quad \forall t. \quad (60b)$$

Clearly, a_t tracks the distance between \mathbf{q}_t and $\boldsymbol{\theta}/\mu$ until the distance becomes smaller than \tilde{B} for the first time; and b_t stops until $\|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| \leq \tilde{B}$ for the first time as well.

With the definitions of a_t and b_t , one can easily show that the recursion (59) implies

$$\mathbb{E} [a_{t+1} | \mathcal{F}_t] \leq a_t - b_t \quad (61)$$

where \mathcal{F}_t is the so-termed sigma algebra measuring the history of two processes. As a_t and b_t are both nonnegative, (61) allows us to apply the super-martingale convergence theorem [24, Theorem E7.4], which almost surely establishes that: (i) the sequence a_t converges to a limit; and (ii) the summation $\sum_{t=1}^{\infty} b_t < \infty$. Note that (ii) implies that $\lim_{t \rightarrow \infty} b_t = 0$, w.p.1. Since $\sqrt{\mu} > 0$, it follows that the indicator function of b_t eventually becomes null and thus

$$\liminf_{t \rightarrow \infty} \|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| \leq \tilde{B}, \quad \text{w.p.1} \quad (62)$$

which establishes that \mathbf{q}_t will eventually visit and then hover around a neighborhood of the reference point $\boldsymbol{\theta}/\mu$.

Proof of (32b) in Theorem 2: In complement to the sample-path result in (62), we next derive (32b), which captures the long-term queue lengths averaged over all sample paths.

Similar to (49), we have

$$\begin{aligned} \|\mathbf{q}_{t+1} - \boldsymbol{\lambda}^*/\mu\|^2 &\leq \\ &\|\mathbf{q}_t - \boldsymbol{\lambda}^*/\mu\|^2 + 2(\mathbf{q}_t - \boldsymbol{\lambda}^*/\mu)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2. \end{aligned} \quad (63)$$

Using the definition $\gamma_t := \hat{\boldsymbol{\lambda}}_t + \mu\mathbf{q}_t - \boldsymbol{\theta}$, (64) can be written as

$$\begin{aligned} \|\mathbf{q}_{t+1} - \boldsymbol{\lambda}^*/\mu\|^2 &\leq \|\mathbf{q}_t - \boldsymbol{\lambda}^*/\mu\|^2 \\ &+ \frac{2}{\mu} (\gamma_t - \boldsymbol{\lambda}^*)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + \frac{2}{\mu} (\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2. \end{aligned} \quad (64)$$

Defining the Lyapunov drift as $\Delta(\mathbf{q}_t) := \frac{1}{2} (\|\mathbf{q}_{t+1} - \boldsymbol{\lambda}^*/\mu\|^2 - \|\mathbf{q}_t - \boldsymbol{\lambda}^*/\mu\|^2)$ and taking expectations on (64) over s_t conditioned on \mathbf{q}_t , we have

$$\begin{aligned} \mu \mathbb{E} [\Delta(\mathbf{q}_t)] &\leq \mathbb{E} \left[(\gamma_t - \boldsymbol{\lambda}^*)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] \\ &\quad + \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] + \mu M^2/2 \\ &\stackrel{(b)}{\leq} \mathcal{D}(\gamma_t) - \mathcal{D}(\boldsymbol{\lambda}^*) + \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] + \mu M^2/2 \end{aligned} \quad (65)$$

where (b) follows from (51).

Summing both sides over $t = 1, \dots, T$, taking expectations over all possible \mathbf{q}_t , and dividing both sides by T , we arrive at

$$\begin{aligned} \frac{\mu}{2T} (\mathbb{E} [\|\mathbf{q}_{T+1} - \boldsymbol{\lambda}^*/\mu\|^2] - \mathbb{E} [\|\mathbf{q}_1 - \boldsymbol{\lambda}^*/\mu\|^2]) &\leq \\ \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\mathcal{D}(\gamma_t)] - \mathcal{D}(\boldsymbol{\lambda}^*) + \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] + \frac{\mu M^2}{2}. \end{aligned} \quad (66)$$

First, it is easy to show that

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{\mu}{2T} (\mathbb{E} [\|\mathbf{q}_{T+1} - \boldsymbol{\lambda}^*/\mu\|^2] - \mathbb{E} [\|\mathbf{q}_1 - \boldsymbol{\lambda}^*/\mu\|^2]) &= \\ \stackrel{(c)}{\geq} - \lim_{T \rightarrow \infty} \frac{\mu}{2T} \mathbb{E} [\|\mathbf{q}_1 - \boldsymbol{\lambda}^*/\mu\|^2] &\stackrel{(d)}{=} 0 \end{aligned} \quad (67)$$

where (c) holds since $\|\mathbf{q}_{T+1} - \boldsymbol{\lambda}^*/\mu\|^2 \geq 0$, and (d) follows from the boundedness of $\|\mathbf{q}_1 - \boldsymbol{\lambda}^*/\mu\|^2$.

We next argue that the following equality holds

$$\lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] = \mathcal{O}(\mu). \quad (68)$$

Rearranging terms in (68) leads to

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\left(\boldsymbol{\theta} - \boldsymbol{\lambda}^* + (\boldsymbol{\lambda}^* - \boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t + \boldsymbol{\theta}) \right)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right]. \end{aligned} \quad (69)$$

Since $\hat{\boldsymbol{\lambda}}_t$ converges to $\boldsymbol{\lambda}^*$ w.p.1 according to Theorem 1, there always exists a finite time T_ρ such that for $t > T_\rho$, which implies that $\|\boldsymbol{\lambda}^* - \boldsymbol{\theta} - (\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\theta})\| \leq \rho$, w.p.1. Hence, together with the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & (\boldsymbol{\lambda}^* - \boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t + \boldsymbol{\theta})^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \\ & \leq \|\boldsymbol{\lambda}^* - \boldsymbol{\theta} - (\hat{\boldsymbol{\lambda}}_t - \boldsymbol{\theta})\| \|\mathbf{A}\mathbf{x}_t + \mathbf{c}_t\| \stackrel{(d)}{\leq} \rho M = \mathcal{O}(\rho) \end{aligned} \quad (70)$$

where (d) follows since $T_\rho < \infty$ and constant M is as in Assumption 4. Plugging (70) into (69), it follows that

$$\begin{aligned} & \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] \\ & \leq \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbb{E} \left[(\boldsymbol{\theta} - \boldsymbol{\lambda}^*)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) \right] + \mathcal{O}(\rho) \\ & \stackrel{(e)}{\leq} \|\boldsymbol{\lambda}^* - \boldsymbol{\theta}\| \cdot \left\| \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbb{E} [-\mathbf{A}\mathbf{x}_t - \mathbf{c}_t] \right\| + \mathcal{O}(\rho) \end{aligned} \quad (71)$$

where (e) simply follows from the Cauchy-Schwarz inequality.

Building upon (59), one can follow the arguments in [14, Theorem 4] to show that there exist constants $D_1 = \Theta(1/\mu)$, and $D_2 = \Theta(\sqrt{\mu})$, for any d , to obtain a large deviation bound as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{P} \left(\|\mathbf{q}_t - \boldsymbol{\theta}/\mu\| > \tilde{B} + d \right) \leq D_1 e^{-D_2 d} \quad (72)$$

where $\tilde{B} = \Theta(1/\sqrt{\mu})$ as in (59). Intuitively speaking, (72) upper bounds the probability that the steady-state \mathbf{q}_t deviates from $\boldsymbol{\theta}/\mu$, and (72) implies that the probability that $q_t^i > \theta/\mu + \tilde{B} + d$, $\forall i$ is exponentially decreasing in $D_2 d$.

Using the large deviation bound in (72), it follows that

$$\begin{aligned} 0 & \stackrel{(f)}{\leq} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [-\mathbf{A}\mathbf{x}_t - \mathbf{c}_t] \\ & \stackrel{(g)}{\leq} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{1} \cdot M \mathbb{P}(\mathbf{q}_t < M) \stackrel{(h)}{\leq} \mathbf{1} \cdot M D_1 e^{-D_2(\boldsymbol{\theta}/\mu - \tilde{B} - M)} \end{aligned} \quad (73)$$

where (f) holds because taking expectation in (72) over all d implies that the expected queue length is finite [cf. (3c)], which implies the necessary condition in (5); (g) follows from [15, Lemma 4] which establishes that negative accumulated service residual $\sum_{t=1}^T \mathbb{E}[\mathbf{A}\mathbf{x}_t + \mathbf{c}_t]$ may happen only when $\mathbf{q}_t < M$ and the maximum value is bounded by $\|\mathbf{A}\mathbf{x}_t + \mathbf{c}_t\| \leq M$ in Assumption 4; and (h) uses the bound in (72) by choosing $d = \boldsymbol{\theta}/\mu - \tilde{B} - M$.

Setting $\boldsymbol{\theta} = \sqrt{\mu} \log^2(\mu)$ in (73), there exists a sufficiently small μ such that $-D_2(\log^2(\mu)/\sqrt{\mu} - \tilde{B} - M) \leq 2 \log(\mu)$. Together with (73) and $D_1 = \Theta(1/\mu)$, the latter implies that

$$\left\| \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \mathbb{E} [-\mathbf{A}\mathbf{x}_t - \mathbf{c}_t] \right\| \leq \|\mathbf{1} \cdot M D_1 \mu^2\| = \mathcal{O}(\mu). \quad (74)$$

Plugging (74) into (71), setting $\rho = \mathcal{O}(\mu)$ in (71), and using $\|\boldsymbol{\lambda}^* - \boldsymbol{\theta}\| = \mathcal{O}(1)$, we arrive at (68).

Letting $T \rightarrow \infty$ in (66), it follows from (67) and (68) that

$$\begin{aligned} 0 & \leq \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathcal{D}(\boldsymbol{\gamma}_t)] - \mathcal{D}(\boldsymbol{\lambda}^*) + \mathcal{O}(\mu) + \frac{\mu M^2}{2} \\ & \stackrel{(h)}{\leq} \mathcal{D} \left(\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\boldsymbol{\gamma}_t] \right) - \mathcal{D}(\boldsymbol{\lambda}^*) + \mathcal{O}(\mu) + \frac{\mu M^2}{2}. \end{aligned} \quad (75)$$

where inequality (h) uses the concavity of the dual function $\mathcal{D}(\boldsymbol{\lambda})$. Defining $\boldsymbol{\varphi} := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\boldsymbol{\gamma}_t]$, and using $\mathcal{D}(\boldsymbol{\lambda}^*) - \mathcal{D}(\boldsymbol{\varphi}) \geq \frac{\epsilon}{2} \|\boldsymbol{\lambda}^* - \boldsymbol{\varphi}\|^2$ in Lemma 1, (75) implies that

$$\|\boldsymbol{\lambda}^* - \boldsymbol{\varphi}\|^2 \leq \frac{2}{\epsilon} \left(\mathcal{D}(\boldsymbol{\lambda}^*) - \mathcal{D}(\boldsymbol{\varphi}) \right) \leq \mathcal{O}(\mu) + \frac{\mu M^2}{\epsilon} \stackrel{(i)}{=} \mathcal{O}(\mu) \quad (76)$$

where (i) follows since constants M and ϵ are independent of μ . From (76), we can further conclude that $\|\boldsymbol{\lambda}^* - \boldsymbol{\varphi}\| = \mathcal{O}(\sqrt{\mu})$.

Recalling the definition $\boldsymbol{\gamma}_t := \hat{\boldsymbol{\lambda}}_t + \mu \mathbf{q}_t - \boldsymbol{\theta}$, we have that

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\frac{\boldsymbol{\gamma}_t}{\mu} \right] - \frac{\boldsymbol{\lambda}^*}{\mu} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[\mathbf{q}_t + \frac{\hat{\boldsymbol{\lambda}}_t}{\mu} \right] - \frac{\boldsymbol{\lambda}^*}{\mu} - \frac{\boldsymbol{\theta}}{\mu} \\ & \stackrel{(j)}{=} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\mathbf{q}_t] - \frac{\boldsymbol{\theta}}{\mu} \stackrel{(k)}{\leq} \frac{1}{\mu} \|\boldsymbol{\lambda}^* - \boldsymbol{\varphi}\| = \mathcal{O} \left(\frac{1}{\sqrt{\mu}} \right) \end{aligned} \quad (77)$$

where (j) follows from the convergence of $\hat{\boldsymbol{\lambda}}_t$ in Theorem 1, and inequality (k) uses the definition of $\boldsymbol{\varphi}$ and $\boldsymbol{\varphi} - \boldsymbol{\lambda}^* \leq \|\boldsymbol{\lambda}^* - \boldsymbol{\varphi}\|$. Recalling that $\boldsymbol{\theta} = \sqrt{\mu} \log^2(\mu)$ in (74) completes the proof.

D. Proof of Theorem 3

Defining the Lyapunov drift as $\Delta(\mathbf{q}_t) := \frac{1}{2}(\|\mathbf{q}_{t+1}\|^2 - \|\mathbf{q}_t\|^2)$, and squaring the queue update, we obtain

$$\begin{aligned} \|\mathbf{q}_{t+1}\|^2 &= \|\mathbf{q}_t\|^2 + 2\mathbf{q}_t^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + \|\mathbf{A}\mathbf{x}_t + \mathbf{c}_t\|^2 \\ & \stackrel{(a)}{\leq} \|\mathbf{q}_t\|^2 + 2\mathbf{q}_t^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + M^2 \end{aligned} \quad (78)$$

where (a) follows from the definition of M in Assumption 4. Multiplying by $\mu/2$ and adding $\Psi_t(\mathbf{x}_t)$, yields

$$\begin{aligned} \mu \Delta(\mathbf{q}_t) + \Psi_t(\mathbf{x}_t) & \leq \Psi_t(\mathbf{x}_t) + \mu \mathbf{q}_t^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + \mu M^2/2 \\ & \stackrel{(b)}{=} \Psi_t(\mathbf{x}_t) + (\boldsymbol{\gamma}_t - \hat{\boldsymbol{\lambda}}_t + \boldsymbol{\theta})^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + \mu M^2/2 \\ & \stackrel{(c)}{=} \mathcal{L}_t(\mathbf{x}_t, \boldsymbol{\gamma}_t) + (\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) + \mu M^2/2 \end{aligned} \quad (79)$$

where (b) uses the definition of $\boldsymbol{\gamma}_t$, and (c) is the definition of the instantaneous Lagrangian. Taking expectations on the both sides of (79) over \mathbf{s}_t conditioned on \mathbf{q}_t , it holds that

$$\begin{aligned} & \mu \mathbb{E} [\Delta(\mathbf{q}_t) | \mathbf{q}_t] + \mathbb{E} [\Psi_t(\mathbf{x}_t) | \mathbf{q}_t] \\ & \stackrel{(d)}{=} \mathcal{D}(\boldsymbol{\gamma}_t) + \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) | \mathbf{q}_t \right] + \mu M^2/2 \\ & \stackrel{(e)}{\leq} \Psi^* + \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t) | \mathbf{q}_t \right] + \mu M^2/2 \end{aligned} \quad (80)$$

where (d) follows from the definition of the dual function (10), while (e) uses the weak duality that $\mathcal{D}(\gamma_t) \leq \tilde{\Psi}^*$, and the fact that $\tilde{\Psi}^* \leq \Psi^*$ (cf. the discussion after (6)).

Taking expectations on both sides of (80) over all possible \mathbf{q}_t , summing over $t = 1, \dots, T$, dividing by T , and letting $T \rightarrow \infty$, we arrive at

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t)] \\ & \stackrel{(f)}{\leq} \Psi^* + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t)] + \frac{\mu M^2}{2} + \lim_{T \rightarrow \infty} \frac{\mu \|\mathbf{q}_1\|^2}{2T} \\ & \stackrel{(g)}{\leq} \Psi^* + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\lambda}}_t)^\top (\mathbf{A}\mathbf{x}_t + \mathbf{c}_t)] + \frac{\mu M^2}{2} \end{aligned} \quad (81)$$

where (f) comes from $\mathbb{E}[\|\mathbf{q}_{T+1}\|^2] \geq 0$, and (g) follows because $\|\mathbf{q}_1\|$ is bounded. One can follow the derivations in (69)–(74) to show (68), which is the second term in the RHS of (81). Therefore, we have from (81) that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\Psi_t(\mathbf{x}_t)] - \Psi^* \leq \mathcal{O}(\mu) + \frac{\mu M^2}{2} = \mathcal{O}(\mu) \quad (82)$$

which completes the proof.

REFERENCES

- [1] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Trans. Automat. Contr.*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.
- [2] S. H. Low and D. E. Lapsley, “Optimization flow control-I: basic algorithm and convergence,” *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–874, Dec. 1999.
- [3] L. Georgiadis, M. Neely, and L. Tassiulas, “Resource allocation and cross-layer control in wireless networks,” *Found. and Trends in Networking*, vol. 1, pp. 1–144, 2006.
- [4] M. J. Neely, “Stochastic network optimization with application to communication and queueing systems,” *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.
- [5] T. Chen, X. Wang, and G. B. Giannakis, “Cooling-aware energy and workload management in data centers via stochastic optimization,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 402–415, Mar. 2016.
- [6] T. Chen, Y. Zhang, X. Wang, and G. B. Giannakis, “Robust workload and energy management for sustainable data centers,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 651–664, Mar. 2016.
- [7] J. Gregoire, X. Qian, E. Frazzoli, A. de La Fortelle, and T. Wongpiromsarn, “Capacity-aware backpressure traffic signal control,” *IEEE Trans. Control of Network Systems*, vol. 2, no. 2, pp. 164–173, June 2015.
- [8] S. Sun, M. Dong, and B. Liang, “Distributed real-time power balancing in renewable-integrated power grids with storage and flexible loads,” *IEEE Trans. Smart Grid*, 2016, to appear.
- [9] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle, “An $\mathcal{O}(1/k)$ gradient method for network resource allocation problems,” *IEEE Trans. Control of Network Systems*, vol. 1, no. 1, pp. 64–73, Mar. 2014.
- [10] J. Liu, A. Eryilmaz, N. B. Shroff, and E. S. Bentley, “Heavy-ball: A new approach to tame delay and convergence in wireless network optimization,” in *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 2016.
- [11] E. Wei, A. Ozdaglar, and A. Jadbabaie, “A distributed Newton method for network utility maximization-I: Algorithm,” *IEEE Trans. Automat. Contr.*, vol. 58, no. 9, pp. 2162–2175, Sep. 2013.
- [12] M. Zargham, A. Ribeiro, and A. Jadbabaie, “Accelerated backpressure algorithm,” *arXiv preprint:1302.1475*, Feb. 2013.
- [13] K. Yuan, B. Ying, and A. H. Sayed, “On the influence of momentum acceleration on online learning,” *arXiv preprint:1603.04136*, Mar. 2016.
- [14] L. Huang and M. J. Neely, “Delay reduction via Lagrange multipliers in stochastic network optimization,” *IEEE Trans. Automat. Contr.*, vol. 56, no. 4, pp. 842–857, Apr. 2011.
- [15] L. Huang, X. Liu, and X. Hao, “The power of online learning in stochastic network optimization,” in *Proc. ACM SIGMETRICS*, vol. 42, no. 1, New York, NY, Jun. 2014, pp. 153–165.
- [16] V. S. Borkar, “Convex analytic methods in markov decision processes,” in *Handbook of Markov decision processes*. Springer, 2002, pp. 347–375.
- [17] R. Urgaonkar, B. Urgaonkar, M. Neely, and A. Sivasubramanian, “Optimal power cost management using stored energy in data centers,” in *Proc. ACM SIGMETRICS*, San Jose, CA, Jun. 2011, pp. 221–232.
- [18] T. Chen, A. G. Marques, and G. B. Giannakis, “DGLB: Distributed stochastic geographical load balancing over cloud networks,” *IEEE Trans. Parallel and Distrib. Syst.*, to appear, 2017.
- [19] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, J. Ramos, and A. J. Caamaño, “Optimal cross-layer resource allocation in cellular networks using channel- and queue-state information,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2789–2807, Jul. 2012.
- [20] A. Ribeiro, “Ergodic stochastic optimization algorithms for wireless communication and networking,” *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6369–6386, Dec. 2010.
- [21] D. P. Bertsekas, A. Nedi, and A. Ozdaglar, *Convex analysis and optimization*. Belmont, MA: Athena Scientific, 2003.
- [22] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena scientific, 1999.
- [23] H. Robbins and S. Monro, “A stochastic approximation method,” *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
- [24] V. K. K. and X. Solo, *Adaptive Signal Processing Algorithms*. Upper Saddle River, NJ: Prentice Hall, 1995.
- [25] N. L. Roux, M. Schmidt, and F. R. Bach, “A stochastic gradient method with an exponential convergence rate for finite training sets,” in *Proc. Advances in Neural Information Processing Systems*, Lake Tahoe, NV, Dec. 2012, pp. 2663–2671.
- [26] A. Defazio, F. Bach, and S. Lacoste-Julien, “SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives,” in *Advances in Neural Info. Process. Syst.*, Montreal, Canada, Dec. 2014, pp. 1646–1654.
- [27] B. T. Polyak, *Introduction to Optimization*. New York, NY: Optimization Software, 1987.
- [28] A. Eryilmaz and R. Srikant, “Joint congestion control, routing, and MAC for stability and fairness in wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1514–1524, Aug. 2006.
- [29] J. C. Duchi, A. Agarwal, M. Johansson, and M. I. Jordan, “Ergodic mirror descent,” *SIAM J. Optimization*, vol. 22, no. 4, pp. 1549–1578, 2012.
- [30] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, “Robust stochastic approximation approach to stochastic programming,” *SIAM J. Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.
- [31] M. Hong and Z.-Q. Luo, “On the linear convergence of the alternating direction method of multipliers,” *Math. Program., Ser. A*, pp. 1–35, 2016.
- [32] H. Karimi, J. Nutini, and M. Schmidt, “Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition,” *arXiv preprint:1608.04636v2*, Oct. 2016.
- [33] Z.-Q. Luo and P. Tseng, “On the convergence rate of dual ascent methods for linearly constrained convex minimization,” *Math. Oper. Res.*, vol. 18, no. 4, pp. 846–867, Nov. 1993.